

AN INVESTIGATION OF THE SUITABILITY OF SMARTPHONE DEVICES FOR ROAD CONDITION ASSESSMENT

By

GUANYU WANG

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

School of Civil engineering
University of Birmingham
November 2018

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

The measurement of road roughness is important for asset management decision making. Not only is road roughness an indicator of road condition and thereby a means of determining road maintenance needs, but it is also used to determine vehicle operating costs (i.e. fuel consumption and vehicle maintenance). Road agencies with large road networks, because of resource issues, are unable to record the condition of the entire network on a sufficiently frequent basis to adequately determine road condition and therefore identify proactive road maintenance requirements. This research investigates whether a smartphone based system may be a suitable means for measuring road roughness at sufficient accuracy and if data from such a system could be used to inform asset management decision making and provide the road user with information about vehicle operating cost of using different routes. This research by means of an in depth review of the literature and the use of a vehicle dynamics package, identified the factors which can most influence the accurate measurement of road roughness by smartphone based systems and quantified the relative importance of these factors. The investigation found that measured vehicle body acceleration, speed, vehicle type and smartphone type are very influential in accurately determining road roughness from a smartphone type approach.

Thereafter, a variety of computational methods were trialled on a multi-variable dataset that had been built using a vehicle dynamic package, to determine if the algorithms could be used to infer road roughness from a dataset which might be available from a smartphone based system. As a result of this analysis, the random forest machine learning algorithm was identified as the most suitable for the task at hand. It was found that the developed algorithm could be used to determine precise measures of road roughness if

data concerning vehicle speed and type, sprung mass, smartphone type and vehicle body acceleration were available. The same algorithm could also be used to classify road condition if only vehicle speed, vehicle type and measured vehicle vertical body acceleration were available in the dataset.

ACKNOWLEDGEMENTS

I thank the University of Birmingham for providing me with the opportunity and facilities to undertake my research.

I would like to express my sincere gratitude to my supervisors Dr Michael P.N. Burrow and Dr G. S. Ghataora; both are senior lecturers in the Department of Civil Engineering, with the School of Engineering at the University of Birmingham. I am appreciative of their time, advice, support and inspiration throughout my research. In particular, thanks are due to Dr Michael Burrow for his invaluable knowledge and experience.

My most sincere thanks go to my parents for their huge support, both financial and mental, and non-stop love that created a comfortable environment in which I was able to complete this research project.

Lastly, I would like to thank all the other members of the department, especially Dr Mehran E. Torbaghan, Research Fellow in the School of Civil Engineering, and Dr Wenda Li for their assistances and advice on various of aspects of the research.

TABLE OF CONTENTS

Chapter 1	1
INTRODUCTION	1
Section 1.1 General	1
Section 1.2 Aim and Objectives.....	3
Section 1.3 Benefits of the Research	3
Section 1.4 Novelty	4
Section 1.5 Outline of the Research	5
Chapter 2	7
ROAD ASSET MANAGEMENT	7
Section 2.1 Introduction.....	7
Section 2.2 Road Asset Management.....	7
Section 2.3 Information and Data.....	11
Section 2.3.1 The World Bank’s Information Quality Level.....	11
Section 2.3.2 Pavement condition	14
Section 2.3.3 Data can be measured using smartphones	22
Section 2.3.4 Methods to relate vehicle body acceleration with roughness	26
Section 2.3.5 Classification of road roughness measuring devices.....	28
Section 2.3.6 Existing smartphone based approaches.....	30
Chapter 3	34
MACHINE LEARNING ALGORITHMS	34

Section 3.1 Introduction	34
Section 3.2 Data mining and machine learning	34
Section 3.2.1 Machine learning algorithms	37
Section 3.2.2 Simple regression analysis	39
Section 3.2.3 The Classification and Regression Tree (CART) algorithm	43
Section 3.2.4 Random Forest (RF).....	47
Section 3.2.5 Gradient-boosted tree.....	51
Section 3.2.6 Neural networks (NNs).....	53
Section 3.2.7 Support Vector Machine (SVM).....	57
Section 3.2.8 K-nearest neighbour (KNN)	59
Section 3.2.9 Naïve Bayes	62
Section 3.2.10 Machine learning algorithms used to determine road condition.....	63
Section 3.2.11 Studies comparing different learning algorithms	65
Chapter 4	71
METHODOLOGY	71
Section 4.1 Introduction.....	71
Section 4.2 Research methodology.....	72
Section 4.2.1 Data requirement for asset management.....	74
Section 4.2.2 Understanding the smartphone	74
Section 4.2.3 Identifying Influencing Factors	75
Section 4.2.4 Developing the Prototype System.....	80

Section 4.3	Summary	83
Chapter 5		85
CASE STUDY: INVESTIGATING THE ACCURACY OF MEASURING ROAD ROUGHNESS USING AN EXSITING SMARTPHONE-BASED SYSTEM.....		85
Section 5.1	Introduction.....	85
Section 5.2	Methodology.....	86
Section 5.2.1	Roadroid Application	87
Section 5.2.2	SCANNER Data.....	88
Section 5.2.3	Experiment Locations	89
Section 5.3	Performance of Roadroid Application	91
Section 5.3.1	Visual Observation	91
Section 5.3.2	Statistical Analysis	95
Section 5.3.3	Threshold Analysis	99
Section 5.3.4	Spectral Analysis.....	103
Section 5.4	Summary	106
Chapter 6		109
FACTORS INFLUENCING THE ASSESSMENT OF ROAD ROUGHNESS		109
Section 6.1	Introduction.....	109
Section 6.2	Characteristics of Smartphones	110
Section 6.2.1	Operating range, resolution and sampling interval.....	111
Section 6.2.2	Accelerometer errors	113

Section 6.2.3 Differences in smartphones when measuring IRI.....	118
Section 6.2.4 Position of the Smartphone.....	121
Section 6.3 Vehicle related factors	126
Section 6.3.1 Vehicle Speed.....	129
Section 6.3.2 Vehicle type.....	131
Section 6.3.3 Suspension stiffness and damping ratio.....	135
Section 6.3.4 Tyre pressure.....	138
Section 6.3.5 Sprung mass	138
Section 6.3.6 Driving style	139
Section 6.3.7 Wheelbase.....	140
Section 6.4 Effects of longitudinal profile	141
Section 6.5 Summary of all factors	145
Section 6.6 Model Development	147
Section 6.7 Discussion	154
Chapter 7	160
SELECTING AN APPROPRIATE PREDICTIVE MODEL.....	160
Section 7.1 Introduction.....	160
Section 7.2 Candidate Algorithms	162
Section 7.3 Building the Algorithms	163
Section 7.4 Performance Evaluation	166
Section 7.4.1 Measures of Accuracy.....	169

Section	7.5 Datasets used for the trials	175
	Section 7.5.1 Data pre-processing	177
Section 7.6	Results of the model selection process	178
	Regression Problem	179
	Section 7.6.1 Measures of all criteria	179
	Section 7.6.2 Weighted scoring system	183
	Classification problem	188
	Section 7.6.3 Measures of Accuracy and Training speed	188
	Section 7.6.4 Other measures of performance	189
Section 7.7	Discussion	192
Section 7.8	Summary	198
Chapter 8		200
	THE DEVELOPMENT OF PROTOTYPE SYSTEM	200
Section 8.1	Introduction	200
Section 8.2	Datasets for developing the computational model	201
	Section 8.2.1 The Variables Considered	201
	Section 8.2.2 Road Profiles	203
	Section 8.2.3 CarSim™ Simulation	203
Section 8.3	Performance of the Computational Model	210
	Section 8.3.1 Regression Model (IQL-2)	211
	Section 8.3.2 Classification model (IQL-3/4)	242

Section 8.4	Summary	249
Chapter 9		251
DISCUSSION		251
Section 9.1	Introduction.....	251
Section 9.2	Research Objectives.....	251
Section 9.3	Limitations of Research.....	253
Section 9.4	Road Friction and Structural Strength.....	254
Section 9.5	Unpredictable Factors Affecting Roughness Measurement	259
Section 9.6	Use of smartphone data to support road asset management	265
Section 9.7	Benefits of the Smartphone-Based System	266
Section 9.8	Additional Considerations	267
Section 9.9	Summary	268
Chapter 10		270
CONCLUSION AND FURTHER RESEARH		270
Section 10.1	Accomplished work	270
Section 10.1.1	Conclusion	271
Section 10.1.2	Findings	272
Section 10.1.3	Remaining problems.....	274
Section 10.2	Future work.....	276
REFERENCES		277
APPENDIX A		294

APPENDIX B.....	297
APPENDIX C.....	337

LISTS OF FIGURES

Figure 2.1: Conceptual components in asset management.....	8
Figure 2.2: Information quality level (IQL) classification	12
Figure 2.3: The specification of wavelengths for different road surface characteristics	18
Figure 2.4: The quarter-car model and other vehicle responses	19
Figure 2.5: The relation between the RMS of acceleration and the square root of the integral of the area under the PSD curve.....	28
Figure 3.1: Steps of the data mining process to gain knowledge	35
Figure 3.2: Typical example of regression	40
Figure 3.3: Binary division in the tree structure for classification (i.e. an example of three classes (1, 2 and 3) decision tree model).....	44
Figure 3.4: A hypothetical plot of the relationship between training and testing error versus tree complexity in the decision tree	46
Figure 3.5: Basic principle of artificial neural networks	54
Figure 3.6: An example of a neural network structure (10-5-1) for the IRI prediction using road defects as predictors.....	55
Figure 3.7: Data converted from 2 dimensions to 3 dimensions using the support vector machine (SVM).....	58
Figure 3.8: An example of K-nearest neighbour classification	60

Figure 4.1:	Components of a road management system.....	71
Figure 4.2:	Research methodology.....	73
Figure 5.1:	Experiment undertaken using Roadroid, A25 in Kent, UK.....	87
Figure 5.2:	Entire route of the experiment on A20, Kent, UK.....	89
Figure 5.3:	Starting point of the experiment on A20, Kent, UK	90
Figure 5.4:	Starting point of the experiment on A20, Kent, UK	90
Figure 5.5:	Comparison of road roughness data measured by SCANNER and those measured by Roadroid (Run One).....	92
Figure 5.6:	Comparison of road roughness data measured by SCANNER and those measured by Roadroid (Run Two)	93
Figure 5.7:	Comparison of road roughness data measured by SCANNER and those measured by Roadroid (Run Three).....	94
Figure 5.8:	Paired sample t-test results of SCANNER data against Roadroid R1 data.	97
Figure 5.9:	Statistical comparison among R1, R2 and R3 datasets.....	99
Figure 5.10:	Comparison of 2m/km threshold analysis (R1 and SCANNER).....	100
Figure 5.11:	Comparison of 3 m/km threshold analysis (R1 and SCANNER).....	102
Figure 5.12:	Spectral density analysis of Roadroid R1 and SCANNER	105
Figure 5.13:	Spectral density analysis of Roadroid R2 and SCANNER	105
Figure 5.14:	Spectral density analysis of Roadroid R3 and SCANNER	106
Figure 6.1:	Common errors of smartphone accelerometers.....	114
Figure 6.2:	Effects of temperature and time on measurements of accelerations using a smartphone.....	118
Figure 6.3:	Car used for testing (left) and mounting position of smartphones (right).....	120

Figure 6.4:	Results of paired sample T-test.....	120
Figure 6.5:	2D stimulation model for passenger cars.....	122
Figure 6.6:	Comparison of signals recoded from two positions of accelerometers mounted in the vehicle	123
Figure 6.7:	Six positions where smartphone devices are normally mounted in the vehicle.....	124
Figure 6.8:	Measured accelerations in y and z axes	125
Figure 6.9:	Factors that influence the vertical accelerations collected using Smartphone data collector.....	127
Figure 6.10:	The effect of vehicle speed on vehicle vertical body accelerations.....	130
Figure 6.11:	Vehicle wheel displacement model for different vehicle driving speeds: a) low speed < 20km/h; b) normal speed (20 to 80km/h); c) high speed, e.g. > 80km/h	131
Figure 6.12:	Simulated Grms of different vehicle types as a function of road condition	132
Figure 6.13:	The effects of adding additional weight on vehicle sprung mass (Speed=50km/h, IRI=3.1 m/km).....	139
Figure 6.14:	Road profiles of two sections (section A and B respectively).....	143
Figure 6.15:	The elevation PSD curves for above mentioned two road profiles respectively.....	144
Figure 7.1:	The selection process to choose the most appropriate algorithm	162
Figure 8.1:	Distribution of datasets in the research.....	205
Figure 8.2:	Overall road conditions of network summarised in the research dataset.....	206
Figure 8.3:	Scatter Plot of IRI against Grms with colours that represent different vehicle types	209
Figure 8.4:	Percentage of the prediction errors between predicted and true IRI using random forest algorithm.....	216

Figure 8.5:	The distribution of four traffic categories in the testing sample.....	219
Figure 8.6:	The effects of traffic composition (i.e. commercial vehicles) on model prediction accuracy	223
Figure 8.7:	The model accuracy of eliminating the vehicle group in turn (4 groups in total)	224
Figure 8.8:	Distribution fittings of simulated Grms of Hatchbacks and Commercial Vehicles	225
Figure 8.9:	Results of Mann-Whitney U test (Group 1: hatchback; Group 2: Commercial vehicles)	226
Figure 8.10:	The histogram plot of testing results (single D-class Sedan fleet)	228
Figure 8.11:	Random error with equal distribution modelled using @Risk.....	231
Figure 8.12:	Frequency plot of the entire original Grms datasets in @Risk.....	232
Figure 8.13:	Frequency plot of the entire modified Grms_variance datasets in @Risk	233
Figure 8.14:	Error of the normal distribution modelled using @Risk	236
Figure 8.15:	Frequency plot of entire modified Grms_variance datasets (with error) in @Risk.....	237
Figure 8.16:	Comparison of Grms before and after introducing a random variable (normal distribution).....	240
Figure 8.17:	Comparison of Grms before and after introducing a random variable (normal distribution).....	241
Figure 8.18:	The process of backward feature elimination loop in the model.....	245
Figure 9.1:	Proposed approach for measuring friction on road.....	257
Figure 9.2:	Vehicle vertical acceleration peak values for concrete bump (M*=average value, S*=standard deviation)	262
Figure 9.3:	Road profile modeified (with a bump), IRI=4.4 m/km	263
Figure 9.4:	Vehicle body acceleration measured on original road profile (IRI=2.4/km)	264

Figure 9.5:	Vehicle body acceleration measured on modified road profile (IRI=4.4m/km)	264
-------------	---	-----

LISTS OF TABLES

Table 2.1:	A summary of management functions.....	9
Table 2.2:	Summary of different information quality levels (IQLs)	13
Table 2.3:	Pavement evaluation and characteristics.....	16
Table 2.4:	The classification of devices used to measure the IRI	29
Table 2.5:	Summary of current smartphone technologies for measuring road defects and roughness	30
Table 3.1:	Data mining tasks.....	36
Table 3.2:	A summary of machine learning algorithms	38
Table 3.3:	The advantages and disadvantages of the decision tree	47
Table 3.4:	Kernels method for SVM.....	58
Table 3.5:	A review of the literature on using machine learning algorithms to determine road condition.....	64
Table 3.6:	Summary of machine learning classifiers	67
Table 5.1:	Satistical analysis of data gathered from Roadroid and SCANNER.....	95
Table 5.2:	Summary of the locations of peak values of IRI measured by two systems	103
Table 5.3:	Root mean square value of measured IRI values from Roadroid and SCANNER	104
Table 6.1:	Vehicle operation speeds required for different smartphone measuring sampling frequencies.....	112
Table 6.2:	Bias measurements of typical smartphone	115

Table 6.3:	Specifications of accelerometers used in shaking table test:.....	116
Table 6.4:	The effects of different smartphone orientations.....	124
Table 6.5:	Effect of angle of smartphone orientation on measured gravity	125
Table 6.6:	Smartphone specification needed to measure road roughness	126
Table 6.7:	Parameters of different vehicle configurations in CarSim™	134
Table 6.8:	The effects of suspension stiffness, damping and tyre pressure.....	137
Table 6.9:	Effects of driving regimes on simulated Grms.....	140
Table 6.10:	The effects of wheelbase in the roughness measurement	141
Table 6.11:	Grms values for four different road profiles.....	145
Table 6.12:	The summarised effects of different influencing factors on the measured vehicle body acceleration	146
Table 6.13:	The values selected for the model development	147
Table 6.14:	Regression Analysis for Large European van and D-class sedan	148
Table 6.15:	Results of machine learning	152
Table 7.1:	Specifications of important parameters in model setting	165
Table 7.2:	Confusion matrix	172
Table 7.3:	Summary of dataset.....	177
Table 7.4:	Normalisation techniques used	178
Table 7.5:	The performances of the algorithms tested to determine precise values of roughness	180
Table 7.6:	Pairwise comparison matrix of the five criteria	183
Table 7.7:	Normalised pairwise comparison matrix for five criteria	184

Table 7.8:	Scores of all algorithms in each measures	185
Table 7.9:	The weighted scores for all algorithms used to obtain precise values of regression.....	186
Table 7.10:	Measures of classification algorithms	188
Table 7.11:	Rankings of classification algorithms	190
Table 7.12:	Weighted scores for classification algorithms.....	191
Table 7.13:	Pairwise comparison matrix of the five criteria in first group	192
Table 7.14:	Pairwise comparison matrix of the five criteria in the second group	193
Table 7.15:	Two groups of weights used	194
Table 7.16:	Weighted scoring for all regression algorithms	195
Table 7.17:	Weighted scoring for all regression algorithm	196
Table 7.18:	Accuracy of random forest algorithm under different model settings	198
Table 8.1:	The variables included and excluded from the simulated dataset	203
Table 8.2:	Summary of the datasets	204
Table 8.3:	Road IRI Classification.....	205
Table 8.4:	A summary of dataset in terms of road condition	206
Table 8.5:	The Pearson's correlation matrix between attributes and outputs	207
Table 8.6:	Sensitivity analysis.....	212
Table 8.7:	Summary of the performance of Random Forest Regression model	213
Table 8.8:	Leave-one-out validation	217
Table 8.9:	Summary of stratified sampling in the vehicle types	218
Table 8.10:	Results of random forest for each category.....	219

Table 8.11: Prediction errors of Vans/Trucks/CP those commercial vehicle types in the computational model	220
Table 8.12: Statistics summary of the simulated Grms (distribution) for Hatchbacks and Commercial vehicles	225
Table 8.13: Statistics of all predicted IRI of testing data (single D-class Sedan)	229
Table 8.14: Compare statistics of original Grms and modified Grms_Variance (with errors) Datasets	234
Table 8.15: Summary of performance of random forest regression using modified dataset.....	234
Table 8.16: Statistics of the original Grms and modified Grms_Variance (normal distribution) datasets.....	238
Table 8.17: Summary of performance of random forest regression using modified dataset.....	238
Table 8.18: Summary of the performance of Random Forest Classification model	243
Table 8.19: the relative important of independent variables	247
Table 8.20: The accuracy of random forest classifier	248
Table 8.21: The Confusion matrix (Test 4)	248

GLOSSARY OF TERMS

Asset The physical infrastructure to be managed

Classification A data modelling task wherein the output is categorical

Construction Development works

Dependent variable Model output(s) where the model is dependent on the given input data

Defect Non fulfillment of a requirement associated with a specific or intended use (e.g. a road condition that has deteriorated and for which repairs have been requested)

FWD Falling weight deflectometer, used to determine a pavement's structural strength

GPS Global positioning system

Grms Root-mean-squared value of vehicle body acceleration in gravity (vertically)

Information Data that is transferred, and made meaningful via processing and dissemination

IQL Information quality level; criteria established by the World Bank, used to group data (based on level of detail) with other features, such as assisting in specifying data collection in a cost-effective manner, and employed in conjunction with road management systems

IRI International Roughness Index

Level of service A subjective measurement of user requirement

Machine learning A method of data analysis using statistical techniques that enable a computer to 'learn' from data and identify unknown patterns automatically without much human intervention, based on the principles of artificial intelligence

Management Coordinated activities to direct and control an organisation by planning and organising the use of resources to achieve a particular goal

Management functions Areas in which road management decisions are made, and which can be divided into planning, programming, preparation and operations management

Network A particular grouping of roads for management purpose (e.g. national road network, trunk road networks, etc.)

Node Fundamental data-processing elements in a tree structure, where each element in the tree structure denotes a test of an attribute that branches into possible outputs

Prototype system An early approximation of a real working system or product, developed for testing concepts and assumptions, and acting as a system that can be replicated or learned from; it provides specifications for a real, complete system rather than a theoretical one

PSD Power spectral density

Rehabilitation Road works that are needed to reinstate a road to a maintainable condition

Road agency A body tasked with carrying out work under an agency agreement (e.g. a road administration)

Road maintenance The group of works that enables a road to continue providing an acceptable level of service; road maintenance can reduce road deterioration, lower road use costs, and keep roads open continuously

Roughness Longitudinal unevenness of a road surface, which impacts on vehicle suspension

Sampling interval A distance or time between which measurements are undertaken or data are captured; with respect to measuring roughness, it refers to the distance between points at which measurements are taken

Screening Initial determination of which road section may need treatment

Strategy A plan or a series of plans that are used to describe a proposed process for implementing policy

Visual inspection A subjective inspection based on simple visual measurement

Chapter 1

INTRODUCTION

Section 1.1 General

In most countries, an efficient road transport system is considered a fundamental prerequisite for general economic and social development and therefore considerable resources are made available for to road construction, network improvements and maintenance. As a result typically road networks have an asset value that represents a significant proportion of national wealth (i.e. gross national product) (Robinson, Danielson and Snaith, 1998). Well-maintained and accessible road networks assist a country's development by facilitating trading, both nationwide and between countries, by reducing transport costs and by improving accessibility to jobs, education, healthcare and recreational facilities. Hence, both the quantity and quality of road networks have significant impacts on all aspects of life (Robinson, 2008).

Due to the socio-economic importance of road networks, they have to be carefully managed, often with scarce resources. The purpose of road management is to maintain and improve the existing road network to the benefit of society, enabling roads to carry traffic in an efficient and safe manner. In other words, road management optimizes the overall performance of the road network over time (Robinson, 2008).

All asset management decisions rely on the availability of appropriate and updated information to evaluate the physical condition, level of service, safety and efficiency of the operation of the road system. However, collecting data can be costly. According to the National Cooperative Highway Research Program Report 334 (McGhee, 2004), the costs

of pavement profile data collection and analysis for agencies ranged between \$2.23 and \$10 per 1.61 km (mile), and the average cost was roughly \$6.12 per mile. For example, merely for the state of Illinois in the United States (i.e. 25th largest state in terms of land area), which has 139,577 miles of roads, the expenditure required to measure the road pavement profile of all roads in the state was approximately \$1.4 million annually (McGhee, 2004).

The large expenditure required and the challenge of processing and updating data are the main obstacles for many road agencies in collecting data on an annual basis. As a result, decisions on maintenance and rehabilitation are often based on outdated data. There is therefore a need to use technology to support asset condition monitoring so that appropriate data can be collected at an appropriate frequency reliably and accurately. One such technology is the smartphone. The smartphone has potential advantages for the collection of large datasets, such as those associated with measuring the condition of road networks, including its potentially low cost, its wide public use and the its regular data update. This thesis investigates such an application. The research postulates the use of a smartphone fitted rigidly to a moving vehicle to measure the vertical accelerations of the vehicle and the subsequent use of the captured acceleration data to support road management.

If such a system can be proved to be achievable, regularly updated road asset management systems that incorporate road condition data collected by smartphones can be used to predict pavement condition, prioritise road maintenance activities on an economic basis and provide road users with information about their costs of using the road network. Accordingly, the research investigates the way in which the physical requirements of measuring road condition using smartphone technology may be achieved and how these

measurements may be affected by a number of influencing factors. Thereafter, a prototype system is developed to demonstrate how data obtained from a smartphone can be manipulated to infer road condition to given levels of accuracy and subsequently how this data can be used to support road asset management decisions.

Section 1.2 Aim and Objectives

This research project investigates if it is possible to use data collected using smartphone technology fitted within moving vehicles to support road asset management decision making. To achieve this aim, the project has the following objectives:

1. To identify from the literature the primary measures of road condition which are needed to make road maintenance decisions, and their associated data requirements.
2. To explore the potential for a smartphone fitted to a moving vehicle to collect the primary measures of road condition as identified in objective one.
3. To investigate smartphone related and other external factors, that may influence the measurement of road condition data collected using a vehicle mounted smartphone.
4. To develop and demonstrate a prototype system that can utilize representative smartphone data to determine to a suitable degree of accuracy, repeatability and reproducibly measures of road condition required to support road asset management.

Section 1.3 Benefits of the Research

The development of a smartphone-based prototype system provides two main benefits to road agencies:

1. Provides the opportunity for the supply of continually updated road condition information
2. Reduces the cost of data collection

Provide the opportunity for the supply of continuously updated road condition

The use of a smartphone-based system to measure road condition has the potential to reduce the cost of data collection for road agencies. A study has found that approximately 63% of the UK population (42.7 million) were smartphone users in 2017 (eMarketer, 2017). The rate of smartphone uptake is still increasing, as is the ability and willingness of smartphone users to utilise associated apps. These factors provide a large market for road condition data collection if road users are encouraged to use the available technology. If the proposed smartphone prototype system is proved achievable, this will enable databases of road condition to be built, which can be updated continuously.

Reduces the cost of data collection

The huge amount of data captured by road users using smartphone technology could provide information on road conditions that will assist road agencies in determining maintenance needs for the road network and in monitoring road conditions. This would transfer resource requirements from road agencies to road users, saving road agencies time and expense for data collection. Furthermore, road users would eventually benefit from such systems due to improved road conditions and decreased road use costs.

Section 1.4 Novelty

The novelty of the research reported herein is as follows:

1. The identification of the most influential factors which might affect the

measurement of road condition using a smartphone, by means of a literature review and a suit of experiments carried out using a vehicle dynamics package allied to computational algorithms (Chapters 2 & 6).

2. A computational model that can be used to predict road roughness from vertical vehicle body acceleration as might be obtained from a smartphone fitted inside a moving vehicle (Chapters 7 & 8)
3. The specification of how accurately road roughness may be predicted as a function of the influencing factors considered (item 1 above) within the computational model (item 2).
4. The identification of the most appropriate classes of vehicles to be used in a system that measures road condition with limited information on other influencing factors (Chapter 8).
5. Identification of a number of measures of road condition which might potentially be obtained using a smartphone (Chapter 9).

Section 1.5 Outline of the Research

To achieve the objectives listed in section 1.2; this thesis is structured as follows:

1. Chapter Two describes, from the literature review, the primary measures of road condition that are required to support asset management decision-making, and whether and how these can be measured via a smartphone mounted in a vehicle moving at normal traffic speeds.
2. Chapter Three reviews potentially suitable computational methods which could be used to interrogate large datasets, such as those associated with road condition data and which might accrue from the suggested smartphone based system.
3. Chapter Four demonstrates the methodology followed in this research to explore

process and techniques required for the development of the system proposed herein.

4. Chapter Six investigates the relative effects of factors that influence the measurement of road vehicle body acceleration using smartphone technology. Data modelling techniques are developed and used to quantify the effects of the above-mentioned factors on the prediction of road roughness.
5. Chapter Five demonstrates a practical application of existed smartphone application.
6. Chapter Seven demonstrates to the process used to select the most appropriate computation algorithms from the list of potentially useful algorithms identified in Chapter Three.
7. The prototype system is presented in Chapter Eight.
8. Chapter Nine presents a discussion of various aspects of the developed system, whilst the conclusion and recommendation for further work are given in Chapter Ten.

Chapter 2

ROAD ASSET MANAGEMENT

Section 2.1 Introduction

This chapter addresses objectives 1 and 2 of the research via a comprehensive review of literature to identify the primary measures of road conditions that are needed to make road asset maintenance decisions, including their associated data requirements. It also explores the potential for having a smartphone fitted to a moving vehicle to collect these primary measures of road conditions.

Section 2.2 Road Asset Management

Roads are assets that require management. Road maintenance management traditionally involves a process in which the pavement system is maintained and monitored via collection and supervision of the current pavement information. Then, future pavement conditions are forecast and maintenance and rehabilitation work are prioritised to achieve a stable pavement system that reaches a predefined performance level (Meegoda and Gao, 2014). However, the process of asset management necessitates that an asset be managed as a business using a combination of management, information, financial, engineering and building practices, among others, that are applied to physical assets to meet economic lifecycle costs. The process is a systematic approach organisations use to optimally manage their physical assets and the relevant performance, risks and expenditures over the assets' lifecycles — achieving the proposed level of service in both short- and long-term planning (Federal Highway Administration and Association of State Highway and Transportation Officials, 1996).

Asset management incorporates business (management and finance), engineering knowledge and economics that are distinguished from infrastructure-focused road maintenance management by providing a service-focused approach, which means asset management combines engineering knowledge and economics to deliver business goals that are generally determined by outcomes or service levels (See **Figure 2.1**).

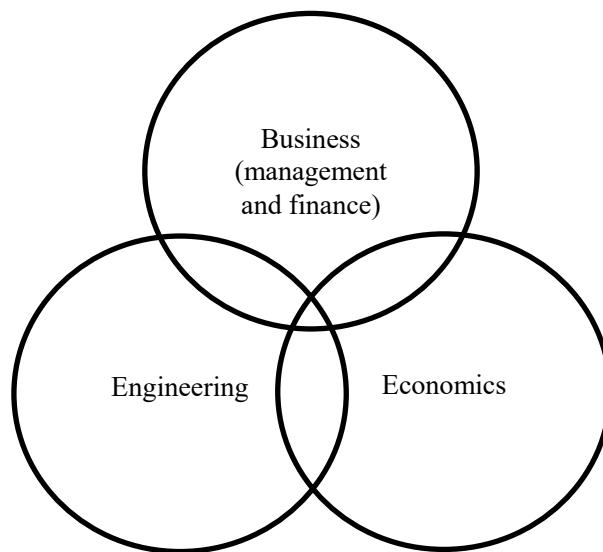


Figure 2.1: Conceptual components in asset management (Robinson, 2008)

An effective road asset management system usually requires data collection methods that are affordable, appropriate, and able to provide relevant information. It also requires a decision support system that is feasible for working with the outputs from different management decisions and strategies (H. Kerali, 2002). This research therefore focuses on investigating how smartphone-based applications can be used to build an information system that allows smartphones to measure road condition accurately to support asset management decisions.

Road management function

Road networks require management decisions to be made, whilst making effective

decisions demands a decision support system. A decision support system comprises powerful processing ability using functions and algorithms that enable sophisticated analysis to be undertaken to predict pavement performance and economic appraisal resulting from asset management decisions (Haas, Hudson and Zaniewski, 1994).

In road asset management, management decisions are made depending on the management functions to be considered. Management functions are generally classified into four distinct levels by which decisions should be made. These functions are usually defined as planning, programming, preparation and operations (See *Table 2.1*).

Table 2.1: *A summary of management functions (developed from (Paterson and Scullion, 1990) & (H. Kerali, Robinson and Paterson, 1998))*

<i>Management function</i>	<i>Time prepared</i>	<i>Responsible Staff</i>	<i>Spatial coverage</i>
Planning	Long-term (strategic)	Senior management and policy level	Network-wide
Programming	Medium-term (tactical)	Mid-level professionals	Network or sub-network
Preparation	Budget year	Junior professionals	Scheme level/sections
Operations	Immediate/very short-term	Technicians/ sub-professionals	Scheme level/sub-sections

Details of each management function are demonstrated as follows (H. Kerali *et al.*, 1998).

Planning: This requires the road system to be analysed as a whole and is particularly useful in long-term preparations (i.e. 5–40 years) or strategic planning estimates of expenditures for road development, preservation and road condition forecasts (key indicators only) under budgetary scenarios. The physical characteristics of road networks are categorised into homogeneous sub-networks by length, traffic level percentages and

pavement structure, and the results of strategic analyses are generally of interest to senior policymakers when planning and considering the economics of the road department.

Programming: This refers to the preparation of multi-year road works and expenditure programmes under budget constraints for sections in the road network that are likely to require maintenance and improvement. The cost-benefit analysis is used in tactical planning to determine the economic feasibility of each set of works. The physical characteristics of the road network are considered on a link-by-link basis, where each link is characterised by road sections and geometric segments depending on its physical attributes. Programming analysis produces prioritised treatment types and estimates of expenditure requirements of treatment type for each section in each year under budget constraints

The main difference between strategic analysis and programming analysis is in how the road links and sections are physically identified. In programming analysis, the links and road sections are identifiable by unique physical units throughout the analysis, whereas in strategic analysis, the road system's individual links and section characteristics are forfeited, as they are grouped into representative, or homogenous sections.

Preparation: This involves preparing short-term planning in which road schemes are packaged for implementation. All roadworks are specified in detail, including bills and quantities, work instructions and contracts. The cost-benefit analysis is revised to confirm the feasibility of the final scheme. Examples of preparation analysis include a detailed design of an overlay scheme and a detailed design for improving road junction and alignment, such as widening lanes and creating additional lanes. In this case, budgets are normally already allocated.

Operation: This covers the activities of road agencies' ongoing operations, daily and weekly. Decisions made include scheduling work to be performed and monitoring labour, equipment and materials. Activities considered typically regard individual road sections or subsections, and the measurements are conducted at a relatively detailed level.

Section 2.3 Information and Data

Managing information is at the heart of road asset management. While understanding the current road system's performance is crucial, being able to predict the effects of proposed policies that could occur in the future is arguably more important. Thus, data are collected to support management decisions in order to 1) determine the optimum road condition, maintenance strategies and expenditure under budget constraints; 2) evaluate the current level of road conditions; 3) determine appropriate levels of investment by prioritising capital improvement and investment in maintenance; 4) simulate the effect of any improvements on future road conditions and the performance of the system; 5) estimate the cost of improvements and control expenditures that occur (Robinson *et al.*, 1998).

Section 2.3.1 The World Bank's Information Quality Level

Making sound decisions requires reliable, updated, affordable and relevant information. Although poor decisions can be made with accurate information, making the correct decision without it is extremely challenging (Paterson and Scullion, 1990).

Data only need to be collected when the need for collection can be justified. Concepts of information quality levels (IQLs) were defined by Paterson and Scullion (1990) that can provide guidance for structuring road information in ways that support different levels of management functions (as shown in **Figure 2.2**). In particular, the IQLs provide a rigorous basis for determining the level of detail that specific data items must achieve to support different levels of management functions (cf, Section 2.2 above).

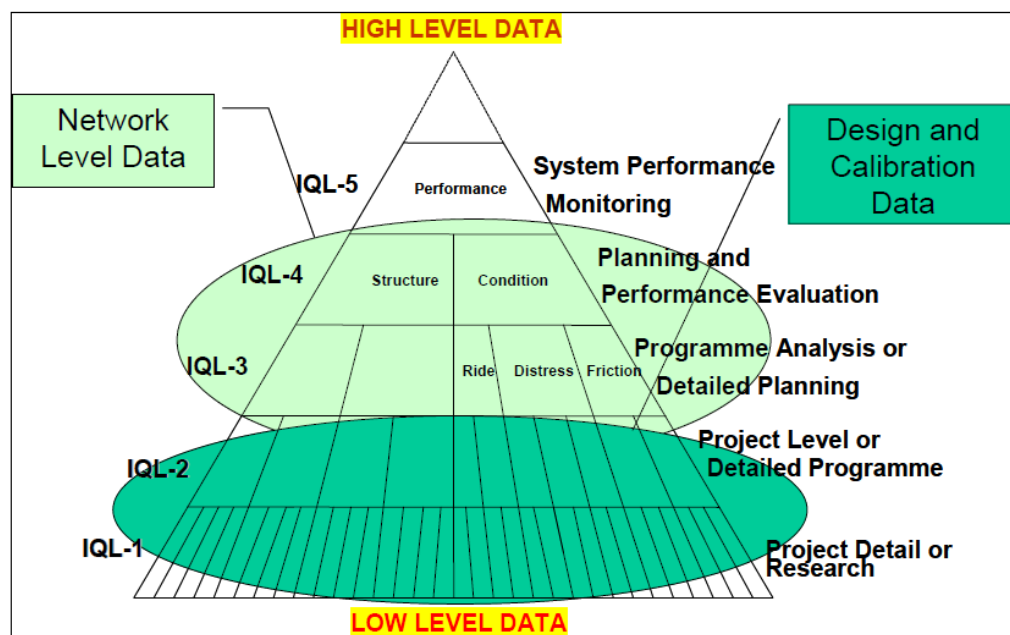


Figure 2.2: Information quality level (IQL) classification (C. R. Bennett and Paterson, 2000)

As can be seen from **Figure 2.2**, a low-level IQL represents very comprehensive information, whereas a high-level IQL means simple, highly summative information that is typically aggregated from very detailed data. More specifically, at IQL-1, the pavement condition has the most comprehensive form with more than 20 attributes that are typically used for project-level decision making. At IQL-2, attributes are reduced from ten to six. At IQL-3, attributes are further decreased to only two or three, namely roughness, surface

distress and skid resistance, which are generally collected for network-level programming. At IQL-4, the pavement condition may be measured by class values, such as good/fair/poor, or a categorical index (0–10). At IQL-5, the road condition is combined with other measures, such as safety, structural condition and traffic congestion, to obtain an indicator that views the highway system as a whole with almost no spatial dimension. IQL-4 and IQL-5 are normally of interest to senior management or the public sector and are presented in a simple understandable form (C. R. Bennett and Paterson, 2000).

To further understand these important concepts, **Table 2.2** provides the detailed IQL examples of the adaption of information to different levels for the road asset management decision support tool HDM-4 (C. Bennett and Greenwood, 2003). HDM-4 is the World Bank's current de facto standard for road investment appraisal, has been used for over two decades to combine technical and economic appraisal of road investment schemes, strategies and standards (H. Kerali *et al.*, 1998).

Table 2.2: Summary of different information quality levels (IQLs)

	<i>IQL-1</i>	<i>IQL-2</i>	<i>IQL-3</i>	<i>IQL-4</i>
Detail of	Most detailed and comprehensive	Detailed	Summary	Most Summary
Information Management function	Operation/ preparation	Operation/ preparation	Preparation/ programming	Strategic planning
Steps in management	Implementation, operation and evaluation	Design and commitment	Feasibility	Identification
Utilisation in HDM-4	Describe the detailed attribute of IQL-2	Actual level of direct inputs in HDM-4	Class-type information for key attributes in each group	Group-level information
Pavement condition	Roughness (m/km), cracking (%), deformation (mm), pothole (%), texture (mm)	Roughness (m/km), cracking (%), deformation (mm), pothole (%), texture (mm)	Riding quality class, surface distress index and friction class	Condition rating by class (good, fair and poor)

Traffic volume	Volume-veh.class 1(veh/day), class 2, ...class <i>n</i>	Volume-veh.class1 (veh/day), class 2, ...class <i>n</i>	AADT(veh/day)	Traffic class (class)
	Volume growth (%/yr)- class 1,...class <i>n</i>	Volume growth (%/yr)- class 1,...class <i>n</i>	Heavy traffic (%AADT) Volume growth (%AADT/yr)	

Section 2.3.2 Pavement condition

Considering road infrastructure and its relevant data collection, different types of data are required for road management. Paterson and Scullion (1990) summarised a list of the information groups used for integrated road management. These groups include road inventory, pavement, structures inventory, traffic, finance, activity and resources. Road inventory is a core file in a road database that identifies the physical component of the road system, consisting of standardised coding and geographic identification of road links and other structures (e.g. bridges and minor features such as lighting, drainage, etc.), as well as all features of environments and climates that are associated with the road system. For the pavement information group (i.e. pavement structure and condition), the pavement condition is regularly measured to monitor pavement deterioration and programme road maintenance and rehabilitation, whereas the pavement structure group is used to forecast major road maintenance expenditures, programmes and designs of road settlement, and rehabilitation and strengthening projects. The traffic information group consists of traffic volume, loadings and accidents, all of which are important pieces of information associated with a variety of uses, especially when traffic loadings causing structural damage that must be considered for structural assessment. Accidents are also useful for understand road safety issues. Finance information is used when decisions and priorities incorporate economic consideration; typical information required includes unit costs,

budget and revenue associated with construction, maintenance, road usage and operation needs. Activity information refers to historic records of projects, such as administrative, technical and financial data. The records of road agencies' property and materials for use in the operational management of a road system are stored in resource information.

Since this research concerns the use of smartphone technology, fitted within a moving vehicle, to collect data for supporting asset management decision-making, emphasis, from among these information groups, has been placed on pavement data because a smartphone accelerometer, by its nature, is capable of measuring a vehicle's dynamic responses to road structure. Other information groups — such as traffic and road inventory — are not considered in the scope of this research.

Road condition information is useful for road users and agencies as it can aid road users in realising vehicle operation costs based on current road conditions, and it can allow them to avoid poor road conditions while driving. Poor road surface conditions, such as roads with potholes, rutting, cracking and surface unevenness, can reduce ride comfort, damage vehicles and increase fuel consumption, which directly increases user costs. Agencies need this information to plan maintenance strategies and programmes (Papagiannakis and Delwar, 2001). Roads must not only be able to provide road users with comfortable, safe and efficient services, but they also must have sufficient structural strength to withstand the combined effects of traffic and environmental loads. To record pavement characteristics that describe performance, both functional and structural data is needed (C. R. Bennett *et al.*, 2007).

Functional Evaluation refers to the evaluation of road surface characteristics that directly affect the safety and comfort of road users and the serviceability of the road. The surveyed

characteristics mainly include roughness, which is associated with ride comfort, as well as skid resistance and surface texture, which relate to safety.

Structural Evaluation is associated with the question of whether the pavement structure is performing satisfactorily under traffic loads and environmental conditions. Surveyed characteristics normally refer to structural performance or mechanistic/structural properties and surface distress.

Table 2.3 summarises the key pavement characteristics considered in the evaluation of pavement condition.

Table 2.3: Pavement evaluation and characteristics (C. R. Bennett *et al.*, 2007)

<i>Evaluation type</i>	<i>Pavement function</i>	<i>Pavement characteristics</i>	<i>Examples of indicators and indexes</i>
Functional evaluation	Serviceability	Roughness	International roughness index (IRI)
			Present serviceability index (PSI)
			Quarter –car index (QI)
	Safety	Texture	Macrotexture
			Microtexture
		Skid resistance	Skid resistance coefficient International friction index (IFI)
Structural evaluation	Structural strength	Mechanical properties	Deflections
		Pavement distress	Cracking
			Surface defects
			Profile deformation

These characteristics are described in detail below.

Roughness

Road roughness is the elevation profile along the travelled wheelpath and is defined as ‘the deviations of a pavement surface from a true planar surface with characteristic dimensions that affect vehicle dynamics, ride quality, dynamic loads, and pavement drainage’ (ASTM,

2015).

Total road roughness comprises a number of components (i.e. different distresses types), and these distresses eventually increase roughness in various ways. For example, surface cracks continue to spall, leading to potholes and, thus, increased roughness. Cracks also allow water to penetrate the road structure, particularly in areas with poor drainage, thus weakening the road. This then leads to structural deformation, or rutting, further increasing roughness. In addition, the magnitude of the effects of these defects on roughness depends on the amount of traffic, pavement materials, maintenance standards, environment and other road conditions (Morosiuk, Riley and Odoki, 2004).

Roughness has been recognised as one of the main criteria used to judge the surface condition of a road network. Road roughness is closely related to serviceability, which has a significant effect on road user costs, safety, comfort and travel speed (C. R. Bennett *et al.*, 2007). Furthermore, road roughness in the frequency domain provides information for comparing the shapes of different road profiles. The mega-texture and short wavelength are primarily associated with superficial pavement defects (i.e. close to the road surface) which affect tyre wear and fuel consumption and thus road user costs. Medium wavelengths (5 to 20 meters) derive mainly from deterioration or deformation in the pavement structure, resulting in poor ride comfort and vehicle damage, whereas long wavelengths (20 to 100 meters), which are normally caused by movements in the foundation or poor structure under the pavement, affect the ride quality only when the amplitude is high (Paterson and Scullion, 1990).

Figure 2.3 demonstrates the typical range of wavelengths for various road surface characteristics. For example, the wavelengths associated with road roughness are generally

between 0.5-50 m. Typically, short wavelengths (less than 3 m) are associated with pavement characteristics in the upper layer, while longer wavelengths (more than 10 m) are related to irregularities in the lower pavement layers (Doré, Flamand and Pascale, 2002).

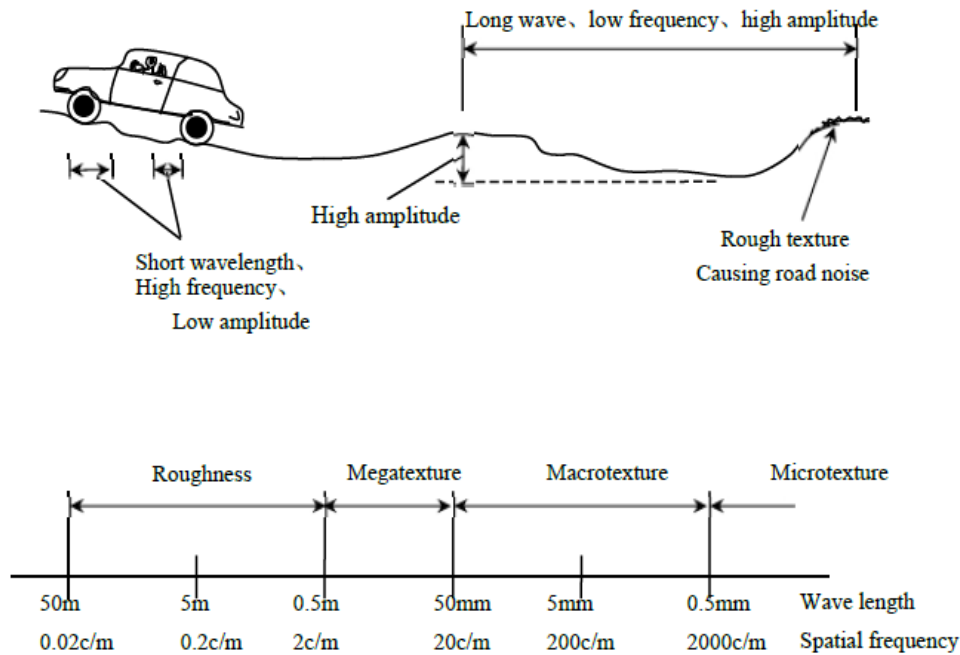


Figure 2.3: The specification of wavelengths for different road surface characteristics (McLean and Ramsay, 1996)

Interest of all road authorities is trying to reduce the roughness to improve the ride quality on the road network since it is a major indication of road condition as well as the road user costs. This therefore leads a number of instruments which claims to measure road roughness, especially those with vehicle-response type system (i.e. measure vibrations received by users in a moving vehicle). However the difficulties in correlating and comparing measures from various instruments into a common scale has demand a time – stable and standardized approach to be utilized. This has led to the International Road Roughness Experiments (IRRE) in Brazil, 1982 (M. Sayers, Gillespie and Queiroz, 1986).

The result from IRRE showed a standard quantitative index that is practical and transferrable, known as the International Roughness Index (IRI). Since then, the IRI has now been widely used and standard measure for evaluating the road roughness (Meegoda and Gao, 2014; Douangphachanh and Oneyama, 2013).

The IRI is essentially a virtual response-type computer-based model based on the response of a mathematical quarter-car model to a road profile. It is calculated by accumulated suspension motions response of the quarter-car model (see **Figure 2.4**), travelling on the road longitudinal profile of interest at a speed of 80 km/h. The IRI is the unit of slope (e.g. m/km).

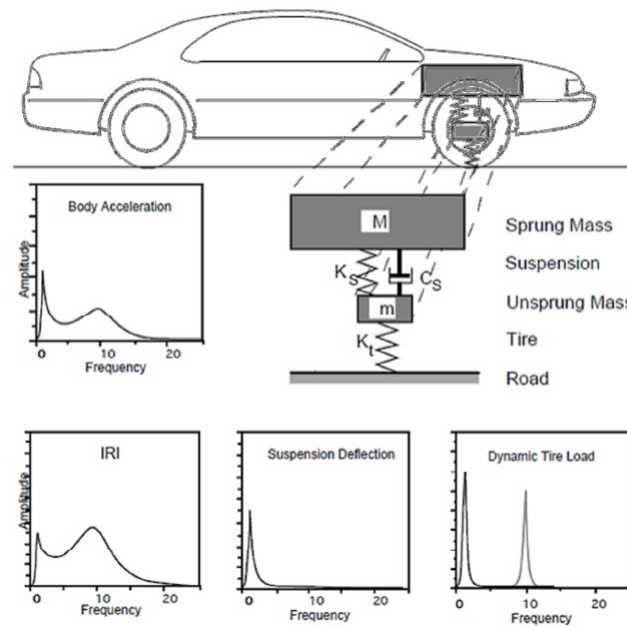


Figure 2.4: The quarter-car model and other vehicle responses (Vitkiena, Nesterovitchb and Puodžiukasc, 2014)

The reason for selecting a quarter-car model is because the IRI is mainly developed to match the responses of passenger car, and the quarter-car is meant to a theoretical

representation of the full car response-type system, the parameters adjusted in the quarter car model is also tried to maximize the correlation with those response-type system (M. W. Sayers and Karamihas, 1998). The IRI is calculated using the following equation (M. W. Sayers, Gillespie and Paterson, 1986).

$$IRI = \frac{1}{l} \int_0^{l/s} |Z_s - Z_u| dt \quad \text{Equation 2.1}$$

IRI is the international roughness index (m/km); l is the length of the road profile in km; s is the simulating speed (80km/h); Z_u is the time derivative of the height of the unsprung mass and Z_s is the time derivative of the elevation of the sprung mass. Typically l is taken to be 160 m.

Pavement texture and skid resistance

These parameters are primarily associated with safety aspects of the roads as they indicate the ability to prevent vehicle skidding. Texture directly affects how well tyres grip on the road, particularly in wet conditions, thus affecting skid resistance. This is because microtexture can penetrate the thin water film between the tyre and the road surface, enabling direct contact, while macrotexture facilitates the drainage of water from the tyre/road contact area (C. R. Bennett *et al.*, 2007). In addition, texture also affects the noise and emission caused by traffic.

Skid resistance is defined as the limiting coefficient of friction between the tyre and road surface and as the ratio of the horizontal frictional force to the vertical force acting on the tyre due to the weight of the vehicle. When the frictional force ‘demand’ exceeds the limiting frictional force, a vehicle skids (C. R. Bennett *et al.*, 2007). From a road management perspective, a road’s skid resistance (e.g. friction) is an important parameter because it can be used as a trigger level for maintenance treatments (e.g. surface dressing

or thin overlay), and it is a function of road safety.

Pavement strength/ mechanical property

The strength of a road structure refers to its ability to support current and future traffic and environmental loadings. Via a structural evaluation, the structural strength should reflect the overall capacity of a road to perform satisfactorily under traffic loads with minimum deformation and distress (C. R. Bennett *et al.*, 2007). When roads carry traffic, the accumulated axle load does structural damage to the pavement layers, and some of this damage can be seen on the surface of the road, such as through potholes or cracks; however, other damage, including road base fatigue and soil settlement, may not be easy to see. Pavement strength is normally estimated indirectly through collecting surface deflection (i.e. deflection bowl) under a dynamic load that is applied to the surface in situ (Robinson *et al.*, 1998). Other mechanical characteristics are also measured for each layer, such as elastic modulus, fatigue properties and resilient tensile stress. A non-destructive method, known as the falling weight reflectometer (FWD), is commonly used to study these characteristics; deflection information is interpreted to determine the pavement's mechanical properties. A widely used approach is to 'back-calculate' the elastic modulus (Von Quintus and Simpson, 2002) and layer thickness of the pavement structure (C. R. Bennett *et al.*, 2007). Because they are used to predict the pavement's remaining life under traffic loads and, therefore, can affect maintenance activities (e.g. rehabilitation), costs and priorities, structural properties are important parameters for making maintenance planning decisions (Paterson and Scullion, 1990).

Surface distress

The appearance and propagation of surface distresses are reflected by deterioration caused

by the combined effects of traffic, environment and ageing. The type, extent and severity of distress directly indicates the structural performance of the pavement and indirectly affect functional conditions (C. R. Bennett *et al.*, 2007).

Surface distress generally includes cracking, potholes, transverse and longitudinal deformation (e.g. rutting) and other miscellaneous defects. Cracking and potholes can vary depending on the pavement type, and they can generally be measured as the percentage of the total measured area or linear units, such as the number of defects. Furthermore, surface deformation (e.g. rutting in the bituminous surface and faulting in the concrete pavement) is usually measured by the vertical deformation of the surface level (C. R. Bennett *et al.*, 2007). These defects are usually measured visually and with more advanced imaging process technology; however, no standardised, similar to IRI, exists for collecting distress data.

Surface distress information is used to control maintenance activities since once these visible defects appear, without timely maintenance activities (i.e. crack sealing and pothole patching), more expensive corrective maintenance (e.g. rehabilitation or reconstruction) will be requested in the future. For example, water can penetrate the road base through unsealed cracks, resulting in structural failure (Robinson *et al.*, 1998).

Section 2.3.3 Data can be measured using smartphones

From the above it may be seen that ideally data collection strategies should consider the road's functional performance (e.g. roughness), structural condition (e.g. deflection, surface distress and rutting) and safety (e.g. skid resistance) (Paterson and Scullion, 1990). Measuring these data necessitates that the principle behind these measurements be understood, which enables this study to determine, in theory, which of these data types can

be measured by fitting a smartphone inside a moving vehicle.

Determining the pavement structural strength involves measures of pavement structural response to an applied load, which is a measure of pavement direct reaction (i.e. surface deflection) to an applied traffic load that causes structural damage (see above). Pavement strength can be measure using destructive and non-destructive methods. Destructive methods require digging a hole so that a sample of pavement thickness and materials in various layers can be taken for detailed analysis, such as determining density, particle size and strength. Non-destructive methods (NDT) such as FWD or Deflectograph are more commonly used than destructive methods for structural evaluation, and most are based on measuring surface deflection under a dynamic, dropped load or a moving, wheel load. NDT and destructive method data are primarily used for project/laboratory-level design and rehabilitation of road sections (i.e. IQL-1/2), whereas, at the network-level (IQL-4), pavement structure data is simply a quantitative class of pavement strength or a classification of remaining life, and more detailed information, such as structural number, can be used at IQL-3.

The measurement of road roughness is effected by two main classes of methods — absolute profile measurement and surface-response (Paterson and Scullion, 1990). Absolute profile measurement directly measures the road's longitudinal profile. Although it is possible to measure profile manually (e.g. with a rod and level or Merlin), in practice, the modern automated method at high operational speed is preferred due to its efficiency and lower labour costs. Some of these devices are inertial or laser profilometers. The surface-response method involves measuring vehicle response to the pavement profile, often using a transducer and accelerometer to measure the distance between the pavement surface and the vehicle. These vertical movements are added to the distance to obtain the

road elevation. Typical examples are the Bump Integrator and the Roadmaster (C. R. Bennett *et al.*, 2007). Generally, the method of absolute profile measurement, representing IQL-1/2, tends to be more accurate than the surface-response approach, representing IQL-2/3. This is demonstrated in the World Bank classification of road roughness measuring devices (See **Table 2.4**).

Surface distress data is usually collected manually by trained surveyors using photographic images to point out the locations and severity of distress on the road (Paterson and Scullion, 1990). However, full automation of distress measurement has become available via hand-held automated recording equipment, photologgers, video image processing and laser detection. Certain types of defects, such as cracking, are still difficult to measure automatically by mechanical means. A number of devices are being increasingly developed in an attempt to measure distress data automatically and comprehensively through integrating laser and image processing techniques into a survey vehicle. However, this application requires high initial costs (i.e. >\$250,000) and vehicle maintenance costs (Paterson and Scullion, 1990). At IQL-3/4, separate ratings or scores should be made for all types of defects, whereas at IQL-1/2, more detailed information is needed for all instances of surface distress, including the type, the amount, the location and the severity level.

For safety considerations, it is important to quantify the ability of a pavement surface to prevent skidding. The measure of skid resistance can be automated using high-speed machines (C. R. Bennett *et al.*, 2007). There are two methods of measurements, which are, a locked-wheel method and a yawed-wheel method. The locked-wheel device (i.e. a wheel that is partially/fully locked to skid) measures braking resistance (i.e. the friction force between the wheel and the pavement) on the longitudinal direction of a wet road surface.

The yawed-wheel method measures the sideways force on a wheel travelling at an angle to the direction of movement. Water is applied to the road surface during the measurement. These devices can be divided into trailer-towed types and vehicle-mounted types. Vehicle-mounted types generally have higher initial costs, and a typical vehicle-mounted device is SCRIM (C. R. Bennett *et al.*, 2007). Similar to surface distress, at the network level (i.e. IQL-3/4), only a quality index is needed whereas, at the project-level, skid resistance should be measured precisely. Pedestrian methods may be needed over short sections to determine the texture depth through manual measurement or a semi-advanced approach (e.g. image processing).

Regarding pavement condition data collection by means of smartphone-based applications, modern smartphones contain inbuilt, three-axle accelerometers and GPS capabilities, which enable vehicle dynamics in the three-axle traveling direction to be captured (Forsl f, 2012). Theoretically, therefore, it may be hypothesised that road roughness can be measured by processing the vertical vehicle body acceleration along the road longitudinal profile (i.e. the surface-response method) at operational vehicle speed and matching these with GPS locations. Surface distress, such as potholes, can be detected by identifying relatively large accelerations from continuous acceleration recordings (Buttlar and Islam, 2014). Cracking may also be measured by processing the recorded images from smartphones (i.e., using the smartphone's in built camera) mounted in any vehicle (Mertz *et al.*, 2014).

For structural strength, it may be possible that a smartphone fitted inside a moving vehicle can indirectly measure the pavement structural response to applied vehicle load. Rut depth or road surface deformation could be estimated via carefully processing, together with roughness, the vehicle's vertical accelerations captured along the road longitudinal profile

in the frequency domain to diagnose pavement condition deterioration trends and remaining functional life at IQL-3/4. Simultaneously, the inbuilt three-axle accelerometer in a smartphone can measure vehicle longitudinal decelerations when braking, which can be used together with information — such as speed, braking distance and the road longitudinal profile — to estimate road skidding resistance.

This research primarily focuses on the road roughness measurement because it is a major determinant of road operational cost, riding comfort and safety for users as discussed above. Road roughness is also an ultimate sign of structural failure in the pavement (Paterson and Scullion, 1990). The measurement of road roughness is very important for managing economic road maintenance. On most road networks, road roughness is usually adopted as the measure of functional condition and its collection can be automated. For measuring other types of defects using smartphones, see the discussion of suitability in [Chapter 9](#).

Section 2.3.4 Methods to relate vehicle body acceleration with roughness

When a vehicle travels on a road, the oscillation input to the vehicle is mainly caused by irregularities of the road surface (Marcondes, Snyder and Singh, 1992). Secondary sources of vehicle dynamic input are factors due to the vehicle itself, including, for example, the vibration and rotation of the vehicle's engine, driveline and tyre/wheel assembly; however, these factors are relatively small compared to those caused by road roughness. It was found that measured vehicle body acceleration on continuous road roughness (i.e. at stable speed) should be regarded as a stationary stochastic process (Chen *et al.*, 2011). Research has recognised power spectral density (PSD) as an appropriate method for describing road characteristics based on the random vibration of the vehicle (Du *et al.*, 2014; Xu *et al.*, 1992; Dodds and Robson, 1973). This is as follows:

Power Spectral Density

The elevation profile of a road, measured over a given length can be seen as a series of sine waves with various amplitudes and phases through the Fourier transform (Davis and Thompson, 2001). The PSD is a plot of these amplitudes against spatial frequencies. Comparing IRI with PSD, the IRI generally provides less information with respect to the shape, or in other words, the characteristics of the measured road profile (Loizos and Plati, 2008), and thus provides information on the distribution of amplitudes of different wavelengths in a road section (T. D. Gillespie, 1992a).

Root-mean-square vehicle body acceleration

A review of the literature suggested that, empirically, a strong correlation exists between the PSD of a vehicle's vertical acceleration and road roughness (Marcondes *et al.*, 1992; Hesami and McManus, 2009; Sun, 2001). However, this process is complex and would require a considerable time to process millions of data pieces that might be expected from a survey of an entire road network. Therefore, an alternative way to use the PSD approach is to use the root-mean-square (RMS) vehicle body acceleration (Grms) in the time series. This method accords with Parseval's Theorem, which demonstrates that the energy of a signal calculated in the time domain or frequency domain should be the same, as shown in Equation 2.2.

$$\int_{-\infty}^{\infty} h^2(t)dt = \int_{-\infty}^{\infty} |H(f)|^2 df \quad \text{Equation 2.2}$$

That is, the integral of the power spectrum, or the square root of the area under the curve, which satisfies the right hand side of Parseval's Equation, should be equal to the left hand of the Equation, which is the RMS value in the time domain (Doertenbach, 2012; Van Baren, 2012; Simmons, 1998; Rogers *et al.*, 1997), See **Figure 2.5**.

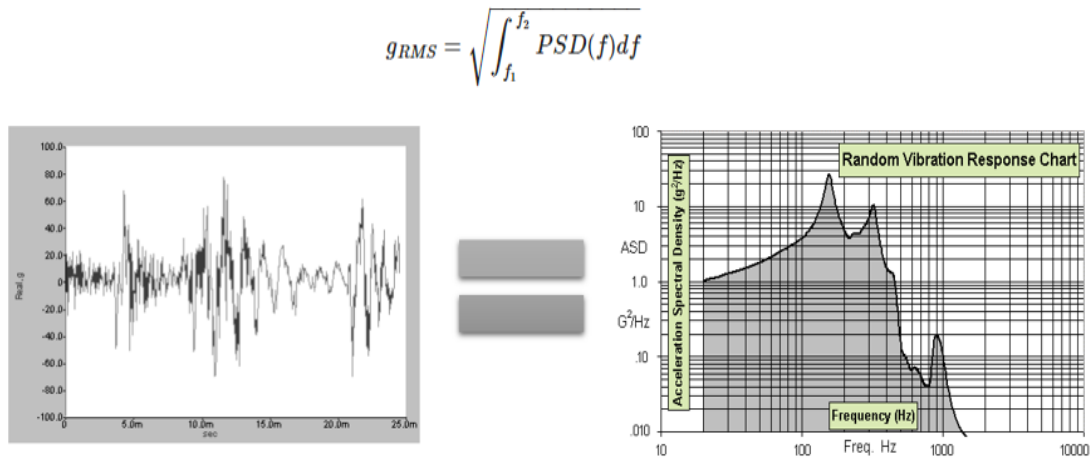


Figure 2.5: The relation between the RMS of acceleration and the square root of the integral of the area under the PSD curve

Furthermore, the RMS approach was also found to correlate well with the IRI (Bridgelall, 2013; Dawkins *et al.*, 2011; Cantisani and Loprencipe, 2010; Marcondes *et al.*, 1992). Considering the use of smartphone-based applications in road asset management, the RMS approach, compared to the PSD approach, appears to have the advantages of simplicity and easy interpretability.

Section 2.3.5 Classification of road roughness measuring devices

Devices for measuring roughness can be categorised into four levels in terms of their accuracy for measuring the IRI. **Table 2.4** shows these classes with their corresponding IQLs, as suggested by the World Bank for road management (M. W. Sayers *et al.*, 1986).

Table 2.4: The classification of devices used to measure the IRI (M. W. Sayers *et al.*, 1986)

	Description	Commercial examples	Details of information	Speciation	Information quality level
Class 1	<ul style="list-style-type: none"> Presents the highest standard of accuracy for measurement of the IRI. Requires that the road longitudinal profile be closely spaced along the travelled wheelpath. Class 1 device is viewed as the primary method to validate other methods. 	Non-contact laser profilers, Face Dipstick, ARRB walking profiler, Rod and level	Most detailed and comprehensive	<ul style="list-style-type: none"> Measurement error of less than 0.3 m/km for paved roads and 0.5 m/km for all other road types; maximum longitudinal sampling intervals ≤ 250 mm Precision of vertical elevation measures between 0.5 mm and 3 mm 	IQL-1
Class 2	<ul style="list-style-type: none"> Profile measurement with an accuracy of less than required in Class 1 due to random and bias errors over a range of conditions. Includes the IRI value computed profiles measured by high-speed profilometers or a static device that does not satisfy the required sampling intervals and precision 	Profilometers (e.g. California, Rainhart); General Motor Research (GMR) inertial profilers	Detailed	<ul style="list-style-type: none"> Measurement error between 0.3 to 0.5 m/km; maximum longitudinal sampling interval > 250 mm and ≤ 500 mm Precision of vertical elevation measures between 1 mm and 6 mm 	IQL-2
Class 3	<ul style="list-style-type: none"> IRI estimated from correlation equations using Response-Type Road Roughness Measurement Systems (RTRRMS) Such devices typically determine roughness using accumulated suspension motion to measure average rectified slope (ABS) or accelerometers as a transducer. Unless an RTRRMS is calibrated by correlation, it cannot be referred to as a Class 3 method. 	Roadmaster; ROMDAS; TRL Bump Integrator; MERLIN; Rolling- Straight Edge	Summary	<ul style="list-style-type: none"> Road condition rated by class (good, fair and poor) or numerically with a measurement error of 0.5 m/km to 1 m/km 	IQL-3
Class 4	<ul style="list-style-type: none"> Subjective evaluation of road condition including the user's perception of ride quality and visual inspection or using measurements from uncalibrated instruments. 	Visual inspection	Most summary	<ul style="list-style-type: none"> Road condition rated by class (good, fair and poor) or numerically with measurement error between 2 m/km to 6 m/km 	IQL-4

The literature suggests that a smartphone technology based road roughness measurement system could potentially satisfy Class 2 or 3 device specifications, according to the road roughness measuring device classification (as demonstrated in **Table 2.4**) (Scholotjes, Visser and Bennett, 2014; Chugh, Bansal and Sofat, 2014; Jones and Forslof, 2014). This means that information collected from smartphone technology for assessing road condition could be suitable for network-level analysis, which generally provides an approximation of roughness used for screening studies and economic analyses for high-level (strategic planning) road management.

Section 2.3.6 Existing smartphone based approaches

According to the literature review, a variety of smartphone-based applications have been developed for measuring road conditions using an empirical approach. **Table 2.5** summarises nine commonly available applications.

Table 2.5: Summary of current smartphone technologies for measuring road defects and roughness

<i>Product</i>	<i>Defect</i>	<i>Description</i>
Street Bump (MIT) (Carrera, Guerin and Thorp, 2013)	Potholes	Web-based crowdsourcing enables confirmation from different users. A spike in acceleration exceeding a threshold is identified and reported as a pothole. The crowdsourcing removes the need for conscious actions on the part of the user, such as shaking or answering a phone.
(Byrne <i>et al.</i> , 2013)	Potholes	Vehicle vertical accelerations are processed using a bandpass filter of 0.5 – 6 Hz to identify and quantify the defects in terms of whether they are major or minor. Spectrum analysis is used to calculate the quotient of a defect on the road. The effects of low vehicle speeds (5km/h), cornering and accelerating/decelerating are eliminated.

'RoadLab' by the World Bank (<i>W. Wang and Guo, 2016</i>)	Road roughness	Uses a regression model to determine IRI from vertical accelerations. The system claims to consider the effects of the position of the smartphone, the vehicle speed and the suspension type for measurement accuracy. However, the tool has not been found to be an accurate predictor of the condition of unpaved roads in fair or poor conditions (Workman, Otto and Irving, 2016).
Roadroid (<i>Forsl�f, 2012</i>)	Road roughness	A multiple linear regression model is used to determine IRI from the root-mean-square of measured vertical accelerations. The model was developed using three vehicle types travelling at speeds ranging from 20 km/h to 100 km/h, and different smartphones were considered. The accuracy of estimated IRI from the correlation was found to be 70%–80% compared to Class 1 (laser-based device), which is considered as IQL-3/4. An appropriate smartphone measuring frequency was suggested to be 100 Hz.
Bump Recorder (<i>Koichi, 2014</i>)	Discrete defects and road roughness	For crowdsourcing, the prototype application was developed using vertical acceleration data obtained from a smartphone fitted on the dashboard of a Toyota Prius. Measurements were made with a frequency of 100Hz. The application claims to determine both roughness and the height of discrete defects by estimating vehicle unsprung elevation from measured vehicle body accelerations using mathematic simulation. The estimated vehicle unsprung elevation is assumed to be equal to the road profile.
Buttlar, W.G. and Islam, M.S. (2014) (<i>Islam et al., 2014</i>)	Road roughness	The system makes use of a linear regression model that relates IRI to vertical acceleration. The model was developed from data captured from a smartphone inside a Honda CRV travelling at 50 mph. The system was found to have high repeatability (coefficient of variance less than 15%). However, the vehicle sprung mass and suspension were found to affect the accuracy.
(<i>Douangphachanh and Oneyama, 2013</i>)	Road roughness	An empirical model relates IRI to vertical acceleration. The data for developing the model was obtained from studies undertaken with two types of smartphones (Samsung Galaxy Note2 and S3) that were fitted inside a Toyota VIGO 4WD pickup truck and a Toyota Camry. The speed of the vehicles was varied, as were the positions of the smartphones inside the vehicles. Measurements were then conducted at smartphone capture frequencies of 100 Hz.
(<i>Du et al., 2014</i>)	Road roughness	A multi-linear regression model was developed between the IRI and the power spectral density (PSD) of measured acceleration data. A Lexus Sedan was used in the experiment, travelling at speeds of up to 60 km/h. The estimated and actual IRI (measured by laser profilometer) were compared, and the error was found to be less than 15%.

<i>(Belzowski and Ekstrom, 2015)</i>	Road roughness	Multi-regression models based on empirical data were obtained from nine different smartphones at a sampling frequency of 100 Hz. Roads of five different classes of IRI values were considered, and better accuracy of prediction was found by categorising road conditions into different classes, instead of predicting the exact IRI. Differences values were found between smartphones when measuring the IRI.
--------------------------------------	----------------	--

As shown in **Table 2.5**, data that are measurable using smartphone-based applications are mainly associated with road roughness and surface distress. A review of the literature has not found representative studies measuring structural strength and surface friction of a road using a smartphone.

Although **Table 2.5** lists a number of available smartphone-based applications, their accuracy of measuring roughness, their usefulness for identifying other defects (e.g. potholes) in reality over large road networks is unclear. Furthermore, the models described in **Table 2.5** are empirical, so they may only be valid for single vehicle fleets. A universal approach that considers a number of many other influencing factors, such as vehicle type and speed, has not yet existed. Thus this research investigates the suitability of using smartphone technology for assessing road condition through a logical process of understanding internal (smartphone-related) and external (vehicle-and profile-related) factors (see **Chapter 6**) and the modelling of vehicle response (see **Chapter 7**). **Table 2.4** was used throughout the research as a guideline to determine the level of accuracy the proposed smartphone-based prototype system can achieve.

Summary

Via desk research, this chapter has sought to understand and describe the primary measures of road conditions that are needed for asset management decision making, including which of these primary measures could potentially be implemented through a smartphone-based application. It was found that, by fitting a smartphone on a moving vehicle, it is possible to measure road roughness, road defects and (potentially) skid resistance and structural strength, which particularly address objectives 1 and 2 in this research. For processing measured vehicle body accelerations, the RMS approach, when compared to the PSD approach, appears to have the advantages of simplicity and easy interpretability. Moreover, this chapter describes a framework for determining the information quality level (IQL) of accuracy that can be achieved by a smartphone-based application, which can support asset management decision making.

Chapter 3

MACHINE LEARNING ALGORITHMS

Section 3.1 Introduction

This chapter has summarised and discussed a number of machine learning algorithms for data modelling in terms of their principles as well as both their advantages and disadvantages. The machine learning algorithms found to be suitable for the task at hand (i.e. correlating the measured vehicle vertical body acceleration to road roughness by considering the influencing variables) will be further used and compared in **Chapter 6** to select the one that is most appropriate.

Section 3.2 Data mining and machine learning

Data mining can be defined as the analysis of an observed, often large, dataset to find hidden relationships within the data that are understandable and useful for data owners (Hand, Mannila and Smyth, 2001). Considering data collection over a road network, using smartphone-based applications entails a large amount of data with many potential predictor variables, such as vehicle types, speeds, weights and suspensions. Therefore, this research considered that the use of machine learning algorithms may be a suitable approach to the analysis of such data sets due to their capability of finding hidden relations in large datasets and handling very complex problems (e.g. non-linearity).

The use of machine learning algorithms to achieve desired goals involves a series of

processes, and the whole process which taken together is often referred to data mining. These processes involve (See **Figure 3.1**): 1) choosing appropriate types of data mining algorithms and how miners organise different algorithms to deal with various problems; 2) pre-processing data (feature selection and data transformation) before data mining; 3) evaluating model performance and presenting findings and after data exploring.

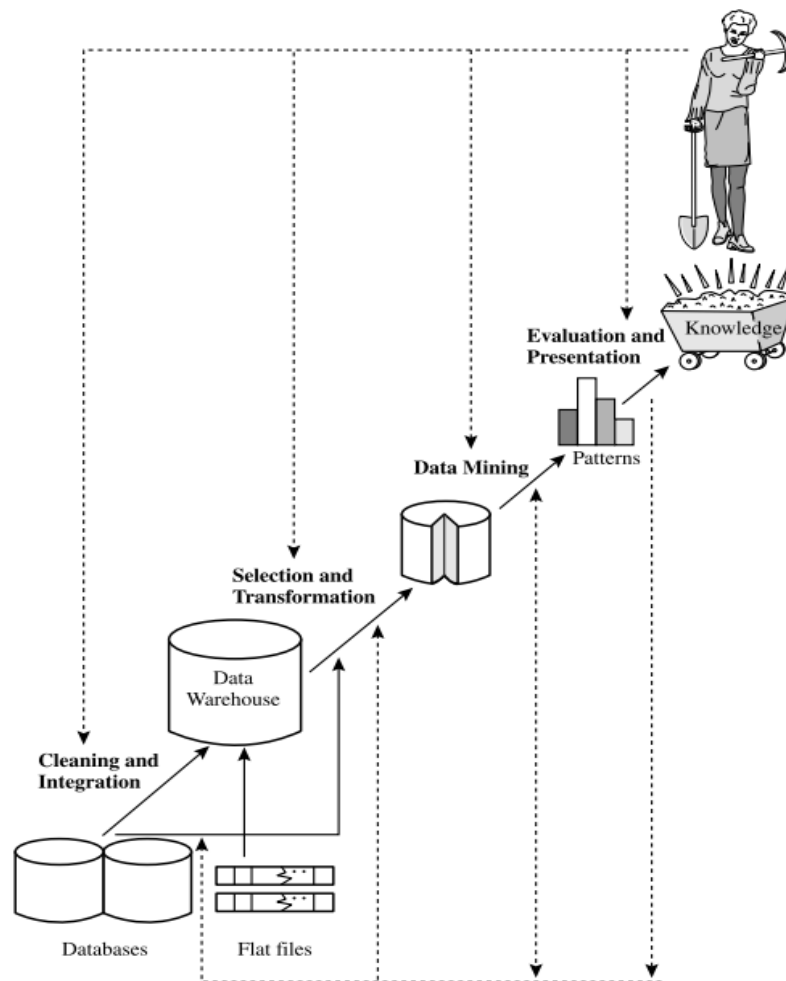


Figure 3.1: Steps of the data mining process to gain knowledge (Han, Pei and Kamber, 2011).

Research has shown that two principles exist that data miners often use to choose an appropriate data mining technique: first, the main goal of the problem and, second, the

structure of the data (Gibert, Sanchez-Marre and Codina, 2010).

Data mining can be broken into a number of tasks as summarised in **Table 3.1** (Hand *et al.*, 2001):

Table 3.1: Data mining tasks

<i>Data mining tasks</i>	<i>Descriptions</i>
Exploratory data analysis	Explore the dataset without any clear ideas about the aims. Methods to use include visualisation and interactive tools (e.g. pie charts, boxplots, and scatter matrices)
Descriptive modelling	Determine the overall probabilistic distribution of the data, through clustering, segmentation analysis and models to describe the relationship between variables by grouping or splitting homogenous data into small sets (i.e. similar data compose small sets).
Predictive model (Regression and Classification)	Build a model that predicts the value of one variable from the known values of other variables, based on historical and current data, to make new predictions on future or unknown events.
Discovering patterns and rules	Recognise patterns, for example, detect fraudulent behaviour in which large data differences are uncovered or find a combination of items that simultaneously or frequently appear within one transaction (association rule).

The primary focus herein is predictive modelling (both regression and classification) to predict road roughness from measured vehicle body acceleration based on historical data, therefore, making new predictions. This prediction process is known as supervised learning. The difference between supervised and non-supervised learning is whether the values of the target/predicted variable are known before data mining. The data structure in predictive modelling is generally divided into qualitative and quantitative. Qualitative

variables refer to categorical variables (i.e. class) — such as gender, status and level — consisting of discrete values. Conversely, quantitative variables refer to numerical values (continuous) and can be any value (Han *et al.*, 2011).

In the data mining process, some machine learning algorithms tend to be more appropriate than those that classify an output into a category (i.e. categorical models). However, a number of machine learning algorithms are compatible with both qualitative and quantitative variables, such as the decision tree (Hand *et al.*, 2001).

Section 3.2.1 Machine learning algorithms

Machine learning explores how computers can learn automatically to recognise complex patterns and make intelligent decisions based on data (i.e. to improve the performance of the model) through computer programs. Many algorithms can deal with predictive tasks, and they are generally categorised into the following two groups based on the data structure used in the prediction (Han *et al.*, 2011):

- Classification
- Regression/numerical prediction

There are a large variety of machine learning algorithms which can fulfil these tasks. To gain a better understanding the types of algorithm which may potentially be suitable, the literature was reviewed in general to appreciate the algorithms which might be suitable. Thereafter a more targeted review of the literature was carried out with a focus on those techniques which have been used in the measurement of road conditions by means of

relating vehicle responses. In addition, a review is presented of specific studies which have compared the advantages and disadvantages of the different types of machine learning algorithms. The findings of these reviews are presented below.

Table 3.2 summarises a number of available machine learning algorithms that may be useful for predictive roughness modelling. These algorithms are described in more detail in the following sections.

Table 3.2: A summary of machine learning algorithms

Algorithm	Tasks	Structure/pattern	Core function	Optimisation methods
Classification and regression tree (CART)	Classification/ regression	Decision tree	Binary split	Greedy search over the structure
Neural network	Classification/ regression	Neuros in network	Backpropagation (non-linear)	Gradient descent
Multi-linear regression	Regression	Multiple variables in linear manner	Linear function	Least square error
Polynomial regression	Regression	Multiple variables in non-linear manner or a special linear regression	High-order polynomial	Least square
Support vector machine (SVM)	Two-class classification/ regression	Hyperplanes/ non-linear boundaries	Linear or non-linear kernel functions for margin maximisation	Dual formulation Sequential minimal optimisation (SMO)
Random forests	Classification/ regression	Ensemble tree	Bootstrap samples	Majority voting /averaging
Gradient-boosted Tree	Classification/ regression	Ensemble tree	Weak learner	Gradient descent
K-nearest neighbour	Classification /regression	Grouping nearest k objects by distance	Euclidean distance	Average of majorities

Section 3.2.2 Simple regression analysis

The purpose of any data analysis is to extract correct estimations from raw information, and to determine whether a statistical relationship exists between independent variables (X) and dependent variables (Y). A common approach is to explore regression analysis. A number of regression models exist that can be used, depending on the distribution of dependent variable Y, and they are often called linear or non-linear. For example, if Y is continued and is approximately normally distributed, a linear regression should be applied, whereas, if Y is dichotomous, a logistic regression is appropriate (Alexopoulos, 2010).

Three common types of simple regression analysis methods — namely linear regression, polynomial regression and logistic regression — are discussed in this section because they are potentially suitable for modelling road roughness. Linear regression is the earliest predictive method that was used to forecast the relationship between predictors X and target variables Y. It is the process of estimating the coefficient of the linear equation, considering one or more independent variables that best fit the model (i.e. continuous dependent variables), and the linear regression model can be written as follows (Rawlings, Pantula and Dickey, 2001):

$$Y = k_1x_i + \varepsilon \quad \text{Equation 3.1}$$

In this equation, k_1 is the coefficient of regression; ε is the intercept; $x_i = 0$; and k_1 is the slope of the line.

It is possible, using regression analysis, to test whether two variables are linearly related

and to test the correlation strength (Menard, 2002). A typical example of linear regression is shown in **Figure 3.2**.

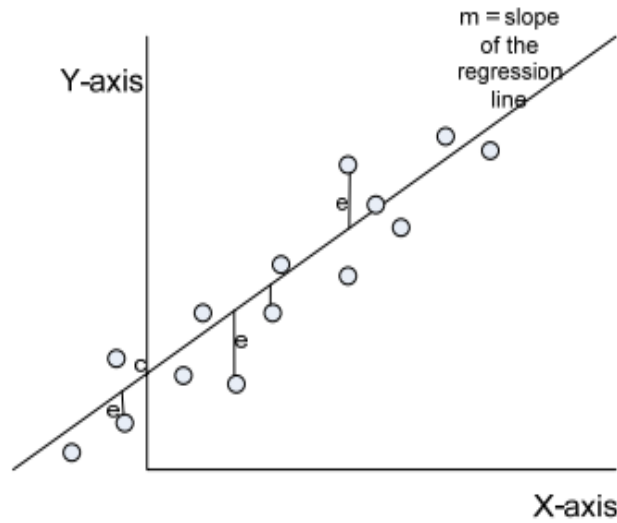


Figure 3.2: Typical example of regression (Nsofor, 2006)

When there are multiple variables in the regression model (i.e., multiple linear regression), the following equation is used (Nsofor, 2006):

$$Y = k_1x_1 + k_2x_2 + k_3x_3 + \cdots \dots + k_nx_n + \varepsilon \quad \text{Equation 3.2}$$

Here, k_n is the coefficient of regression, and ε is the error term.

Multi-variable linear regression is often used as the road roughness modelling method based on measured vehicle responses (Durst *et al.*, 2011; Douangphachanh and Oneyama, 2013; Cantisani and Loprencipe, 2010; Du *et al.*, 2014), where the IRI is often referred to as target Y, and predictor variables are generally considered to be the vehicle speeds and measured vehicle body accelerations. This technique is simple to understand, and it is

considered the most basic method that engineers try first. Although modelling the IRI using linear regression has been commonly adopted, it suffers significantly in terms of accuracy when the data is collected from traffic consisting of more than one vehicle type, travelling at a range of speeds (Abdelaziz *et al.*, 2018; Kargah-Ostadi, 2014).

In addition, there are a number of assumptions associated with the use of linear regression, including (Lewis-Beck and Lewis-Beck, 2015):

1. A linear relationship is assumed to exist between input and output
2. There is no interaction or strong correlation between predictor variables
3. The error term ε is random (uncorrelated) and normally distributed with zero mean and constant variance, homoscedasticity
4. The error term is independent
5. No or few outliers and variables should be of a known functional form
6. The predictors variables are not correlated

These assumptions must be met before performing the linear regression to achieve the best unbiased linear estimators. If these assumptions are partially met, the model is ill-conditioned. For example, if multicollinearity exists between independent variables, the parameters estimated for collinear variables may produce high standard errors (Washington, Karlaftis and Mannering, 2010). Standard linear regression is not an appropriate algorithm, especially for road roughness predictions, due to correlated predictors and its inability to cope efficiently with non-linear relationships (Nitsche *et al.*, 2012; Kargah-Ostadi, 2014).

Sometimes, linear regression may not fit the data satisfactorily. Therefore a more complex, higher degree polynomial regression is often explored. The polynomial regression is considered to be a special form of linear regression in which the relationship between the dependent variables, Y , and predictor variables, X , is modelled as polynomial functions (Peckov, 2012).

The polynomial regression has commonly been used in the literature for empirical modelling of road roughness due to its simplicity and flexibility in shape (Chandra *et al.*, 2012; Marcondes *et al.*, 1992). However, some limitations were found with this method. Polynomial regression has poor interpretability. Particularly, high-degree polynomials lead to a model with a complicated structure. The high-degree polynomials may fit the model well within the data range, while the performance drops rapidly outside the range of data (another distribution of the dataset), indicating poor generalisation ability (Peckov, 2012).

The assumptions mentioned above are likely to be unmet if the dependent variable has only two or three categories. In particular, a binary dependent variable, where Y is categorised as 1 or 0, may be problematic for applying ordinary least square linear regression due to nonlinearity; thus, the maximum likelihood logistic regression can be used (Menard, 2002). In logistic regression, the interest is whether the classification of dependent variables into one or another category can be predicted by independent variables. Thus, instead of predicting the arbitrary value associated with the dependent variables, the problem has been reconceptualised as predicting the probability that a case

falls into one category or another (Menard, 2002).

The logistic regression groups the data into classes using the linear decision boundary known as the linear classifier, and the logistic regression has been used, with reasonable accuracy, for modelling the probability of asphalt concrete cracking severity (Y. Wang, 2012) and road surface condition (Hassan, 2015). Its advantages, such as requiring limited computation power and allowing relatively quick training speeds, makes this approach attractive to many researchers; however, the linearity boundary affects its performance in solving complex classification problems (Razi and Athappilly, 2005). In addition, the logistic regression can perform well on small datasets rather than dealing with large datasets (Perlich, Provost and Simonoff, 2003).

Section 3.2.3 The Classification and Regression Tree (CART) algorithm

The initial CART model was developed based on its binary division of a dataset using tree structure which enables it to readily deal with complex datasets containing a number of variables in both regression and classification problems (Breiman *et al.*, 1984). The difference between classification and regression trees is that with the regression tree, the dependent variable Y is continuous (i.e. real-valued), whereas with the classification tree, the dependent variable Y is categorical. The CART models are commonly used due to their advantages as non-parametric (i.e. no hypothesis is needed) methods. That is, no underlying relationship is required between the dependent and independent variables (Chang and Chen, 2005). With a tree structure, the data is split (i.e. binary division) by

answering nodular questions in each tree node (i.e. yes or no). The root node (the first) contains the entire sample, while the remaining tree nodes consist of a subset of the dataset. **Figure 3.3** illustrates an example of a small-scale classification tree structure (i.e. three classes), and the classification tree's binary split can act as a linear boundary for the best segmentation of data plots.

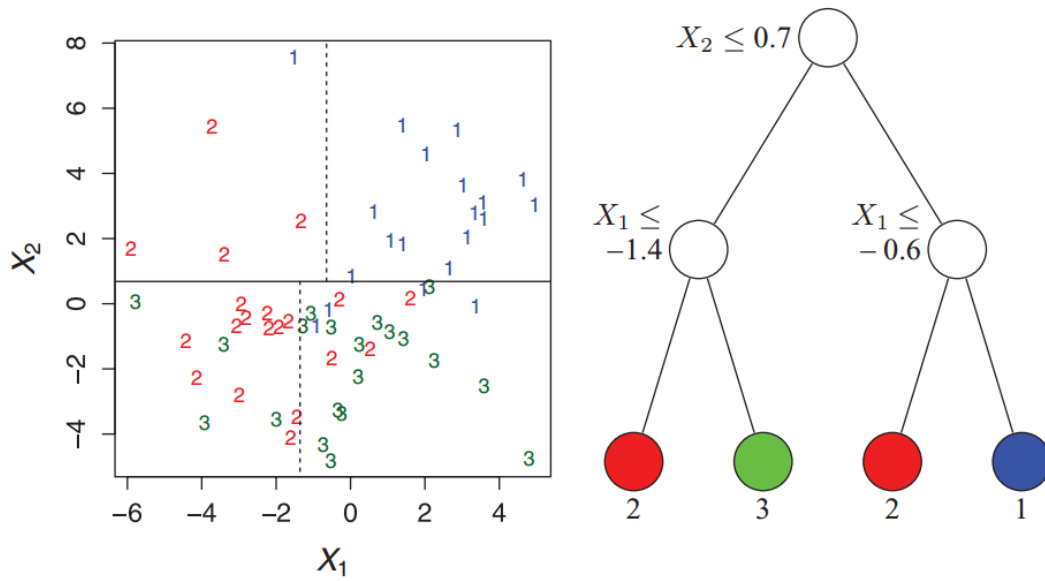


Figure 3.3: Binary division in the tree structure for classification (i.e. an example of three classes (1, 2 and 3) decision tree model) (Loh, 2011)

The greedy algorithms are chosen to grow a tree using all possible independent variables until it archives to the maximal tree size. The best split of predictor variables in each division generally has the steepest descent in terms of the greedy algorithm (i.e. mostly improves the performance) (Chou, 1991). This process is undertaken by searching all values in the independent variables to find the best partition (i.e. minimise the weighted divergence) of the parent nodes into the child nodes. That is, in each split, the dataset is

split into two subsets that are mutually exclusive, attempting to keep each subset as homogenous as possible (De'ath and Fabricius, 2000; Chipman, George and McCulloch, 1998). The distance between data in the two groups of child nodes separated by this binary division is maximised (Dasgupta *et al.*, 2011). Thereafter, a new prediction is made by searching for a unique terminal node from the root node until reaching the bottom of the tree.

The goal of the decision tree is to split the dataset into homogenous subsets and keep the size of the tree relatively small. However, this may not be easily achieved because if splitting is continued, the tree can grow constantly until no further split can be made or until only one class is left in the node. In this case, it is vulnerable to over-fitting the data because decreased testing error is generally associated with a more complex tree structure. The extra nodes that have been added to the tree structure merely predict the noise of the data (i.e. unpredictable data with large variance).

Figure 3.4 demonstrates this issue in detail via a hypothetical plot of the relationship between training and testing error versus tree complexity.

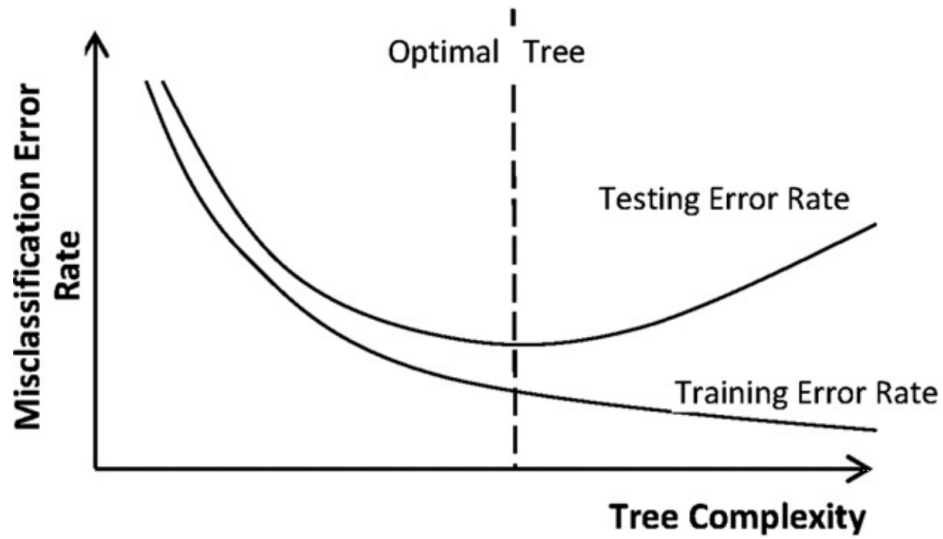


Figure 3.4: *A hypothetical plot of the relationship between training and testing error versus tree complexity in the decision tree (Kashani and Mohaymany, 2011)*

Figure 3.4 shows that although the training error rate decreases with increased tree complexity; the testing error rate, on the contrary, appears to increase at the same level of model complexity after achieving the optimal number of trees. This is because the model has already over-fit the training dataset, therefore, when a different dataset obtained from the same population (i.e. with a different distribution) has been used for testing, the developed model does not perform equally well.

In addition, the decision tree also has a number of advantages and disadvantages. For example, it is easy to interpret, even though developed trees can be very complex. Another advantage is that it is relatively fast to train (Dasgupta *et al.*, 2011). It has also been suggested that CART performs adequately in small datasets, but is less competitive in handling large-scale problems when compared to neural networks (NNs) (Markham, Mathieu and Wray, 2000). However, decision tree models run much faster than NNs with a shorter learning curve (Danjuma, 2015). In addition, it has also been found that the

decision tree is unstable, indicating that any change in the training dataset can largely modify the splitting rules in the tree structure (Kashani and Mohaymany, 2011; Chang and Wang, 2006). Therefore, success may depend on the available data (Han *et al.*, 2011). The advantages and disadvantages of using the decision tree are summarised in **Table 3.3** (Schlotjes, 2013).

Table 3.3: The advantages and disadvantages of the decision tree

Advantages	Disadvantages
<ul style="list-style-type: none"> • Suitable for small datasets • Good interpretability • Trains faster than neural networks • Comparable to neural networks • Performs better than linear regression 	<ul style="list-style-type: none"> • Tends to overfit the dataset • Easily affected by any changes made in the training dataset • Not appropriate for very large-scale problems because it generates extremely complex tree structures

The CART models were widely used in the study of road safety which are complex and involves considerable predictor variables (Nayak *et al.*, 2011; Chang and Wang, 2006; Kashani and Mohaymany, 2011). For example, the CART model was used to determine the factors that contributed to road crashes in Queensland. This model achieved an accuracy of 95% compared to NNs (83%), multiple linear regression (70%) and SVM (68%) (Emerson, Nayak and Weligamage, 2011). Then, the non-parametric CART model was selected to investigate the severity of accident injuries in Taiwan. Results suggested that the CART model can perform well in predicting injury severity (90.3%) (Chang and Wang, 2006).

Section 3.2.4 Random Forest (RF)

Random forest (RF) (Breiman, 2001) is an extension of the decision tree, and it consists of

an ensemble of many decision trees that fit into different bootstrap samples of the training subsets, which is also called tree bagging. The high variability can be achieved by bagging all trees (Friedman and Hall, 2007; Breiman, 1996). RF is also called ensemble learning; more than one tree is combined and grown in a new way. In ensemble learning, a complex problem is divided into many sub-problems that are much easier to solve. Ensemble trees have been found to provide improved robustness and accuracy over single models (García-Pedrajas, Hervás-Martínez and Ortiz-Boyer, 2005). Each tree in the RF is grown in association with a random data subset in which final prediction is acquired by aggregating the ensemble (Biau, 2012). It is built based on the CART but fits the data better due to its bootstrap sampling and the majority voting rule utilised for prediction with a large number of bagged decision trees (typically around 500 decision trees) (Berk, 2005).

This entire algorithm can be summarised as follows (Liaw and Wiener, 2002):

1. A number of bootstrap samples (N_{tree}) are drawn from the original dataset.
2. An un-pruned classification or regression tree is grown for each bootstrap sample while different principles are applied to the trees. At each node, instead of choosing the best split among all predictors, a random selection (M_{try}) of predictors is used, and the best split is chosen from those variables.
3. New predictions are computed by aggregating the results from each N_{tree} using the majority voting rule for classification and average for regression.

4. For each bootstrap iteration, the data not in the bootstrap samples — which is called ‘out-of-bag’ (OOB) data — is used to predict the error.
5. The prediction errors are aggregated using OOB data, and the average error rate is computed.

The OOB samples are used for computing the internal estimation of the strength and correlation of the trees (roughly 36.8%). They are necessary because, if the predicted responses for observations are based on the same data set used for building such a model, overoptimistic results can be measured (Berk, 2005). In addition, in the RFs, each tree is grown at the maximal depth with no pruning. A random subset of the original dataset (bootstrap sample) is used for training, and for each split, only a random number of subsets of predictor variables are selected.

RF can provide low bias and low variance estimations due to their randomness during bootstrap sampling and predictor variable selection, meaning they force diversity between the base learners (trees) (Zhang and Haghani, 2015). Thus, RF works well with very large datasets (Dasgupta *et al.*, 2011). The advantages of RF can be summarised as follows (Reynolds *et al.*, 2016; Kouzani, 2007):

1. Very accurate and able to deal with a large number of examples
2. Able to measure the relative importance of each predictor variable
3. Very efficient compared to some other machine learning classifiers
4. Free of normal distribution assumptions

5. Robust to outliers and data noise
6. Learn fast due to parallel computing

Bootstrap samples from the dataset, which help reduce the risks of overfitting, are an alternative way of picking independent, random samples from a well-defined population (Berk, 2005). A major issue related to RF is that there are so many trees to interpret, always leading to the problem of less interpretability (Biau, 2012).

RFs are commonly utilised in road image processing (Kaczalek and Borkowski, 2016; Emery and Singh, 2013). They are also used in road surface identification due to their good generalisation ability (i.e. do not limited to some specific datasets) and fast computing time (Fernández, Huerta and Prati, 2015). Research based on sensor data collected by buses has also showed that RF perform very well on multi-class, road type classifications, and they are also recommended for road sign identification systems because of their ability to cope with complex images with huge variances (Kouzani, 2007). Nitsche et al. (2012) conducted a study using vehicle responses to evaluate road roughness. The RF algorithm was found to perform well.

Nevertheless, no clear guidance exists on how to determine the optimal parameters used in RFs, which typically include the number of trees (N_{tree}). The optimal number of trees increases with the number of predictors, and the best method involves comparing a performance of the forest using a full dataset with the performance of the same forest using only a subset of the dataset. If they perform equally well, the number of trees is sufficient.

Section 3.2.5 Gradient-boosted tree

Boosting (Kearns and Valiant, 1988), a machine learning method, is based on the strategy that determining many general rules is much easier than finding a highly accurate, single rule for prediction. Boosting begins with finding a basic rule using a method or algorithm. Then, different training data subsets (different distributions over the data sample) are fed into it iteratively. For each iteration, a new prediction rule is generated, and, subsequently, after many iterations, many weak rules are combined into a single prediction rule, which is possibly much more accurate than any one of the preceding weak rules (Schapire, 2003). Theoretically, boosting can be applied to any algorithm, with the original idea being to boost the accuracy of many weak learners, for example the decision tree (Breiman, 2001). A detailed process of boosting algorithms is as follows (Zhang and Haghani, 2015).

1. Given a sample distribution of D , determine the total number of base learner M
2. Define the initial distribution of the training sample as $D_1 = D$
3. Loop: for $m = 1$ to M :

Train a base learner $B_m(X)$ using the training sample distribution D_m

Computer the error of the learner

Adjust the distribution D_m to D_{m+1} , and train the learner again to make the mistake of the learner more distinct

Output the constructed base learner $B_m(X)$

End loop

4. Output the prediction of the tree ensemble for a new input X : $\frac{1}{M} \sum_{m=1}^M B_m(X)$

Gradient boost (Friedman, 2001), rooted in Adaboost (Freund and Schapire, 1996), has a different optimisation approach known as gradient descent. It is a non-parametric algorithm in boost learning with a decision tree that is commonly used as a weak learner. This method combines the advantages of the tree-based model with the strength of the boosting algorithm. One distinct advantage of this approach is that it can handle many predictors in both regression and classification tasks without any pre-processing or transformation. Additionally, it is insensitive to outliers and can perform well with partially inaccurate data because it can exclude the irrelative input variables. However, the gradient-boosted tree has been found to be unstable when constructing large trees (Ahmed and Abdel-Aty, 2013).

In addition, the gradient-boosted tree is functionally close to the RF because of their tree like features and the characteristic of randomisation during tree construction (i.e. randomly selected subsets to improve the algorithm). However, RF actually build trees in parallel, and the majority vote is applied to the prediction; the gradient boost actually creates a series of trees, and the performance of predictions gradually improves by adding more trees, which is known as the step-wise manner (Olinsky, Kennedy and Brayton Kennedy, 2014). Each step of adding one more learner is aimed at correcting the mistakes generated by previous learners, more emphasis is placed on incorrect predictions in the boosting algorithm because those samples that are misclassified or incorrectly predicted have a higher chance of being selected with higher weightings. Weightings and parameters of the previous learners remain unchanged until the additional learners are added to minimise the

lost function (e.g. least square error in regression). Lost function measures the difference between predicted and true values, and the gradient descent procedure is applied to ensure the best convergence of the lost function (Zhang and Haghani, 2015). In this case, the performance of the model increases with the number of trees that are added.

However, being overly fit the training data generally leads to poor generalisation when the model becomes extremely complex with many iterations and deep tree depth, again increasing the risk of over-fitting. Thus, the optimal number of iterations and the tree depth are needed when constructing such a model to minimise the risk associated with prediction.

Section 3.2.6 Neural networks (NNs)

A NN model (Ripley, 1996) contains many features, including nonlinearity, massive parallelism and processing by multi-layer neurons, as well as dynamic feedback among units (Kuan and White, 1994). One important characteristic is its ability to learn from examples and create its own rules. NNs can be trained to perform specific functions by adjusting the weights of connection between neurons. Thus, inputs lead to specific outputs, and this process is undertaken by comparing the predicted outputs and targets (Demuth and Beale, 2001). This process is known as back-propagation, as shown in *Figure 3.5*.

Figure 3.5: Basic principle of artificial neural networks (Demuth and Beale, 2001)

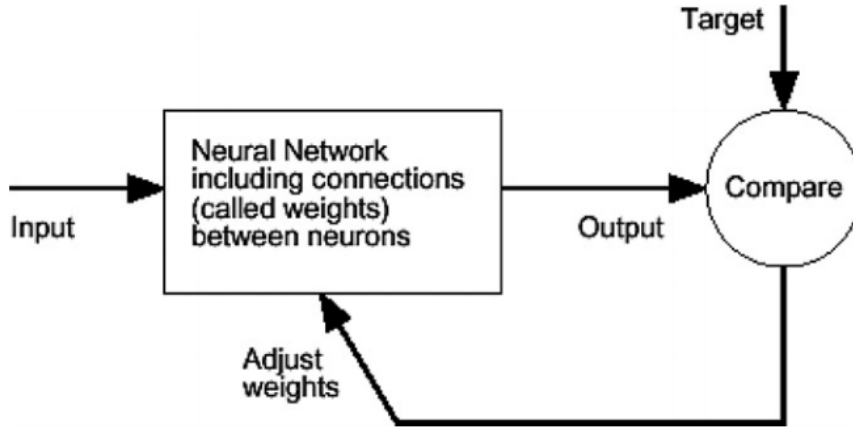


Figure 3.6 illustrates an example of the structure of the NN structure for predicting road IRI. The basic units composing an NN is called the process elements (PE), or, simply, neurons. The PE only process a single output from all the information they receive, and the single output is propagated down the connections. In the case of road IRI prediction, many predictor variables are included, such as condition, traffic, structural and environmental variables. These input variables can be seen as vector $X = (a_1, a_2, \dots, a_n)$, and an associated parameter with each input is called the weight (well known as connecting strength) coming from vector $W_{ij} = (W_{1j}, \dots, W_{ij}, \dots, W_{nj})$, in which these weights present the connection strength from PEa_i to PEb_j . A bias term, θ_j , is used to compute the predicted output value b through W_{0j} , which is the final weight computed by the output value b (Attoh-Okine, 1999):

$$b_j = f(\sum a_i W_{ij} - W_{0j}\theta_j) \quad \text{Equation 3.3}$$

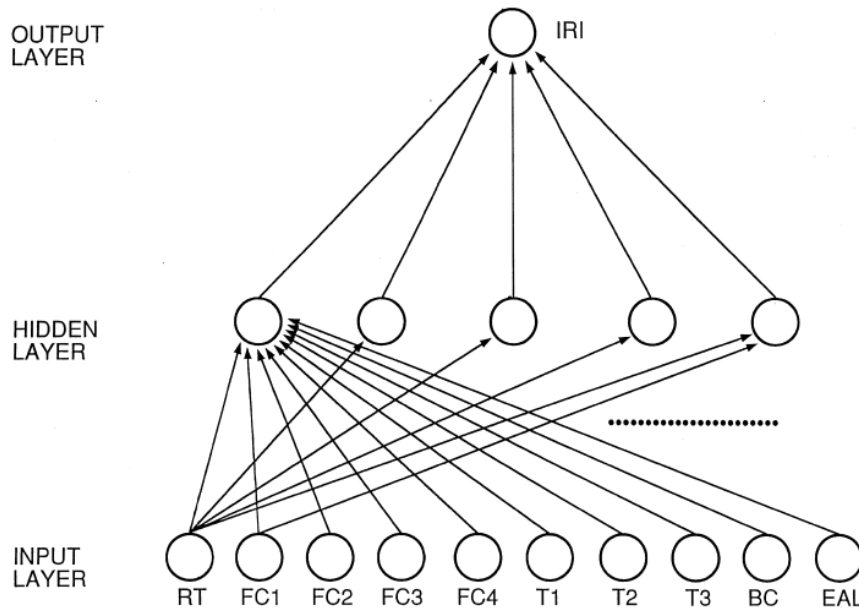


Figure 3.6: An example of a neural network structure (10-5-1) for the IRI prediction using road defects as predictors (Attoh-Okine, 1999)

A NN model is typically used when sufficient data samples are available for both training and testing, and it is also known for having a high capability of coping with ‘noise’ and complexity in the dataset, even if there is no prior knowledge of the forms of the functional relationship between predictors and target variables (Razi and Athappilly, 2005). Typically the NN is regarded as being highly accuracy; however, it also has its own drawbacks due to its black-box structure (i.e. only viewable in terms of its inputs and outputs without any knowledge about internal workings). The interpretability of NN models is always mentioned. The nature of the NN structure prevents the interpretation of results and stops insights into the reasons that model performance changes due to different micro-architectural components (ElMoustapha *et al.*, 2007). Thus, obtaining a reasonable

interpretation of the entire structure of networks is difficult.

However, the nature of NN structure also makes the algorithm quite suitable for predicting road conditions (Attoh-Okine, 1994). NN possess a ubiquitous, non-linear feature with no assumptions respecting the underlying distribution in the data (Nitsche *et al.*, 2012). Furthermore, NNs can make generalisations and offer real-time solutions for very complex problems due to its strong interconnections and massive parallelism.

An NN model was successfully used to estimate the present serviceability index of road pavements (i.e. a commonly used road functional condition indicator for ride quality) by Terzi (2007). This model was also compared with the regression model, and the NN outperformed it (with the R square equal to 0.85) (Terzi, 2007). Attoh-Okine (1994) conducted a study using a back-propagation NN model to predict road roughness where structural deformation (i.e. a combination of structure number, accumulated traffic, surface defects, cracking, layer thickness and rut depth), environmental issues and other non-traffic mechanisms were considered the predictor variables. The results suggested a strong correlation between the predicted and actual IRI. In addition, Attoh-Okine (1994) suggested that NN is suitable for such a task because they are taught to explore the data and allow learning, while also being told the appropriate responses to different inputs.

Still, there is no clear guidance for choosing how many layers and neurons should be used in an NN. Theoretically, the NN model, which contains one hidden layer with sufficient neurons, can accommodate any continuous function (Hruschka, 1993; Tsoukalas and Uhrig,

1996; Bailey and Thompson, 1990). Results have suggested that a learning rate between 0.2 and 0.5 provides strong performance (Refenes, Zapranis and Francis, 1994).

Section 3.2.7 Support Vector Machine (SVM)

The main idea behind SVM is to transform the space of input variables to a high-dimensional space via linear or non-linear functions defined by inner product functions (Yan *et al.*, 2011). Therefore, the model turn the samples (i.e. in two-dimensions) into mapped points in space, and, thus, the examples of the separate categories are easily divided by a clear gap (i.e. hyperplane) in the space, as demonstrated in **Figure 3.7**. The SVM generates an optimal separating hyperplane in a high-dimensional space that gives the least error and maximum margin. That is, the distance between the hyperplane to the neighbouring points is maximised. A new example is then mapped into that same space and predicted to belong to a category based on the part of the hyperplane in which it is located. This inner function is often called the kernel function. **Table 3.4** summarises different kernel functions and their associated boundary shapes.

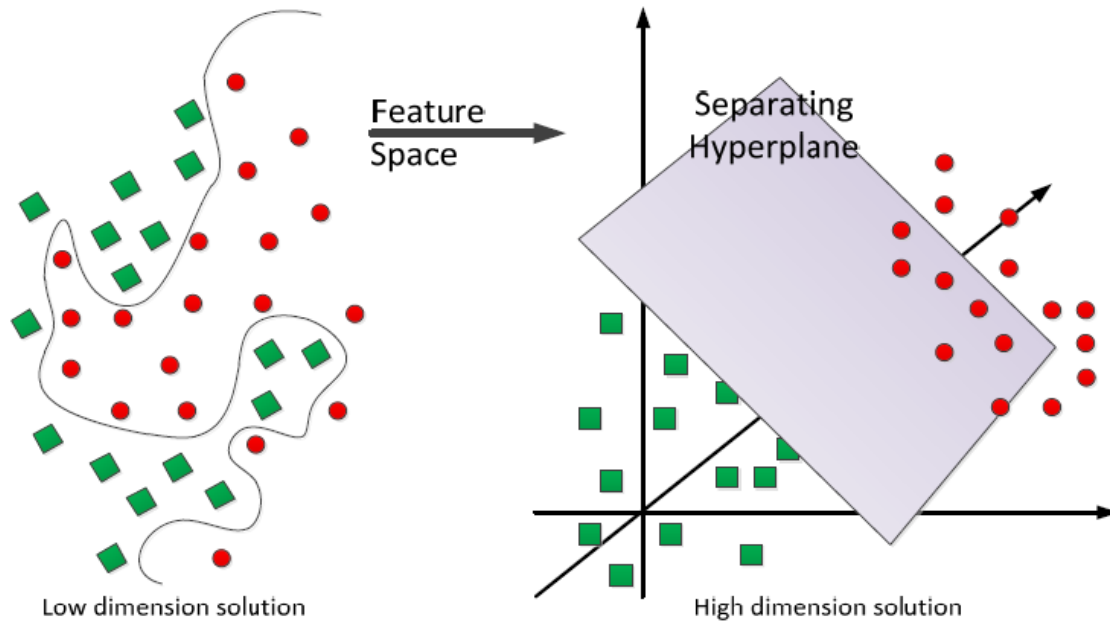


Figure 3.7: Data converted from 2 dimensions to 3 dimensions using the support vector machine (SVM) (Schlotjes, 2013)

Table 3.4: Kernels method for SVM (Schlotjes, 2013)

<i>Kernel Method</i>	<i>Description/Decision boundary</i>	<i>Equation</i>
Linear	Straight line, through the data	$k(x_n, x_m) = x_n^T x_m$
Polynomial (of degree d)	Curve, through the data	$k(x_n, x_m) = (1 + x_n^T x_m)^d$
Radial Basis Function (or Gaussian)	Circular, surrounding the data	$k(x_n, x_m) = e^{-\gamma(x_n - x_m)^T(x_n - x_m)}$
Sigmoid	S-curve, through the data	$k(x_n, x_m) = \tanh(\gamma x_n^T x_m + c)$

The SVM was initially developed for binary classification problems, which separate the dataset into positive and negative classes. Due to its distinct advantages in supporting non-linear boundaries between classes, also called support vectors, the SVM has been

commonly used for many classification and regression problems. With suitable kernel functions, the classification problem becomes simple in the feature space.

The SVM model is robust against over-fitting; however, it is also referred to as a black-box structure method, which is similar to NN (ElMoustapha *et al.*, 2007). The SVM model has also been found to be suitable for generalising and solving many realistic problems with small sample sizes (Yan *et al.*, 2011). For example, a least-square support vector regression (SVR) in MatlabTM exists and supports only up to 20,000 rows of examples (Pelckmans *et al.*, 2002). In addition, training the SVM model is particularly slow. According to ElMoustapha *et al.* (2007), they spent 10 hours training an SVM model, whereas they only spent less than 10 minutes training a model decision tree based on the same data set.

The SVR (Vapnik, Golowich and Smola, 1997) is commonly used in successfully predicting traffic accidents (Wu, Chen and Zheng, 2011) and traffic flow (Rong and Chun, 2010; Liu, Li and Shao, 2007). Yan *et al.* (2011) investigated the use of the SVR model to successfully estimate pavement serviceability ratings. The performance of the SVR model was compared to the regression model, and the results suggested that the SVR model performs better than the regression model in terms of the root-mean-square error and R^2 .

Section 3.2.8 K-nearest neighbour (KNN)

The idea underlying the nearest neighbour is perhaps the most straightforward out of all the machine-learning algorithms. In the classification of the problem, testing data are classified based on the classes of their nearest neighbours (Cunningham and Delany, 2007).

This is also known as example-based classification, or a lazy learning algorithm, because, during the run-time, all examples are required and stored in memory for performing new classifications (Cunningham and Delany, 2007).

The KNN method can be used for both classification and regression problems. It uses the data directly, without building a model first. Thus, the details of the model construction are not required. Merely one parameter, K , is defined, and number of nearest neighbours should be included to predict class membership (Dreiseitl and Ohno-Machado, 2002). The main idea of KNN is that, if the closest K examples belong to certain a class (i.e. majority vote), then the sample is assigned to this class.

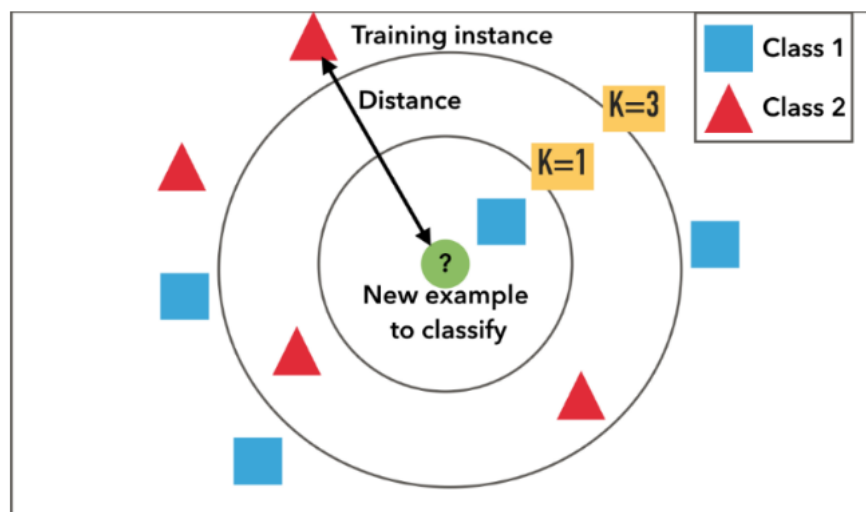


Figure 3.8: *An example of K-nearest neighbour classification (SFL Scientific, 2015)*

The principle of the majority vote is applied to a group of K -nearest neighbours to define the appropriate class label. **Figure 3.8** demonstrates how the majority vote is used via an example. As shown in **Figure 3.8**, if $K = 1$, the new example is indicated as a blue square

(Class 1), whereas, if $K=3$, the new example is indicated as a red triangle, as the majority of the nearest neighbours are red triangles (two out of three). By varying K , the model becomes less flexible (small K value) or more flexible (large K value) (Dreiseitl and Ohno-Machado, 2002). If the K value is small, the local estimate behaviour can be poor due to data sparseness and noisy, mislabelled points, whereas a large K value betters the estimation while smoothing outliers from other classes that are included. Selecting an appropriate K value and distance metric can largely influence the performance of the KNN classifier (Imandoust and Bolandraftar, 2013).

The advantages of KNN are simplicity, robustness to noise and effectiveness when the training dataset is large. However, algorithm simplicity often leads to performance deficiency and inappropriate usage in complex problems. Additionally, users only define the distances between nearest neighbours, and this model can misrepresent the problem when relationships between data are complex or unknown (Dreiseitl and Ohno-Machado, 2002). KNN is also sensitive to irrelevant or redundant attributes because they contribute to the similarity in finding neighbours (Imandoust and Bolandraftar, 2013).

KNN regression simply takes the average of the output variable of K -nearest examples. This method is often chosen in studies of, for example, traffic speed (Cai *et al.*, 2015) and traffic flow analysis (Yoon and Chang, 2014; Zou *et al.*, 2015; Yu *et al.*, 2016). The forecasting model is based on historical databases, and it performs well due to advantages of a non-parametric approach. KNN does not need deep knowledge to understand

modelling due to the simple model parameters, which is beneficial when time and budget are limited.

Section 3.2.9 Naïve Bayes

The Naïve Bayes is a statistical classifier that enables prediction of the probability that a given tuple will fall into a particular class (Han *et al.*, 2011). It is based on Bayes' theorem, which assumes the value of an attribute (predictor variable) on a given class is irrelevant to (i.e. independent from) the values of other attributes. The assumption is to simplify the computations involved in the Bayes classifier, which are often regarded as naïve (Han *et al.*, 2011). An example was used to demonstrate how the probability can be estimated, which is described as follows. Assume that data examples are confined to customers described, respectively, by the attributes (independent variable) of age and income. Data example X is a customer who is 35 years old with an income of \$40,000.

Suppose that H is the hypothesis that this customer will purchase a computer, and $P(H/X)$ is the posterior probability that customer X will purchase a computer when the age and income of this customer are provided. In contrast, $P(H)$ is the prior probability, whereas in this example, it is the probability that any given customer will purchase a computer irrespective of their age or income and other information (i.e. independent of X). Additionally, the probability $P(X/H)$, known as the posterior probability of X conditioned on H, is the probability that customer X is 35 with an income of \$40,000, given that we already know this customer will purchase a computer. The $P(X)$, the prior probability of

X, is the probability that a person from our list of customers is 35 years old with an income of \$40,000. These probabilities follow Bayes' theorem, which also provides the following equation for calculating the posterior probability $P(H/X)$ (Han *et al.*, 2011):

$$P(H/X) = \frac{P(X/H)P(H)}{P(X)} \quad \text{Equation 3.4}$$

Despite its naïve design and oversimplified assumptions, the Naïve Bayes classifier works well in solving a complex problem. It has advantages, such as fast training, and it is easy to interpret and can cope with discrete and continuous attributes (Han *et al.*, 2011). Research has shown that the Naïve Bayes can perform equally well with the decision tree in many classification problems (Doreswamy, 2012; Han *et al.*, 2011), and only a small amount of data is required to train the Naïve Bayes classification (Doreswamy, 2012). Despite its advantages, the Naïve Bayes is generally believed to perform less well than other machine learning algorithms (e.g. NN) due to its assumption of conditional independence (Bhavsar and Ganatra, 2012).

Nevertheless, the Naïve Bayes was used to successfully recognise road surface (i.e., differentiating between dry and moist) and road roughness based on reflection intensities from laser scanner measuring on a road (Aki *et al.*, 2016). The accuracy achieved by Aki *et al.*, 2016 in their study was 92%.

Section 3.2.10 Machine learning algorithms used to determine road condition

Summary of machine learning algorithms used to determine road condition from vehicle based measurements. **Table 3.5** below summarises a review of the use of machine learning

algorithms for the measurement of road conditions by means of relating vehicle responses.

Table 3.5: A review of the literature on using machine learning algorithms to determine road condition

<i>Literature</i>	<i>Principles</i>	<i>ModelUsed</i>	<i>Reason chosen</i>
(Zeng <i>et al.</i> , 2014)	Road condition classification (deficient or not) through vehicle acceleration RMS value.	Logistical regression	Feasible for binary classes classification (accuracy greater than 92%)
(Nitsche <i>et al.</i> , 2012)	Road roughness prediction based on vehicle responses	1. ANN 2. SVM 3. Random forest	Does not state, however, selected algorithms were commonly used in the reviewed literature (Root Mean Square Error <0.3 m/km)
(Ward and Iagnemma, 2009)	Terrain (road type) classification algorithm using unsprung mass (wheel) accelerations	Support Vector Machine (SVM)	Computational capability (correct estimation of the surface type of 89.4%)
(DuPont <i>et al.</i> , 2005)	Terrain (road type) classification using sensor data from the robotic vehicle	Probabilistic neural network (ANN)	Simplicity, robustness to noise and fast training speed (accuracy of greater than 85%)
(Soleimani and sahebi, 2012)	Predicts road roughness using vehicle vertical wheel acceleration	Backpropagation neural networks (NN)	Capable of solving highly non-linear and complex problems Efficient for processing imprecise and noisy data
(Rajamohan, Gannu and Rajan, 2015)	Road condition and surface type classification using three-axle acceleration data and image data	K-Nearest Neighbour (K-NN) Performs well	Not stated
(Seyfi <i>et al.</i> , 2013)	Road condition evaluation using fast speed laser-based deflection measurement (TSD)	Decision tree (regression)	Popular predictive models Capable of handling complex and non-linear relationships High accuracy and comprehensive model that is understandable

Section 3.2.11 Studies comparing different learning algorithms

A comparative study of different machine learning algorithms for various regression and classification problems was conducted in this research. The findings, summarised in this section, are used to support the algorithm selection process described in **Chapter 6**.

Regression

Nitsche et al. (2012) conducted a comparative study between NN, the SVM and RF for road roughness predictions using vehicle responses. The NN were selected as a reference for comparison due to their advantages of computational capacity in solving non-linear problems as a non-parametric approach. A total of 1,000 testing datasets were used consisting of 35 variables, and the results suggested that the three models perform similarly in terms of the mean absolute error between predicted and actual value; however, the SVM was more stable than NNs in handling large numbers of variables. Nevertheless, a comparison of NN and SVR for the road IRI prediction in another study suggested that NN performs better than SVR, with greater accuracy (R-squared= 0.94, RMS Error = 0.99 m/km) and with a smaller standard deviation of mean squared error (Shahnazari *et al.*, 2012).

A comparative study of different machine learning algorithms was undertaken for real estate data, particularly the number of influencing factors affecting house prices. The results of this study showed that NN, the SVM and RF are the best performers (Santibanez, Kloft and Lakes, 2014). The effectiveness of machine learning regression algorithms on

computer performance analysis was investigated, and four common algorithms were compared, including the multi-linear regression, NNs, decision trees and SVM, and the results showed that the NNs and decision tree outperformed others (ElMoustapha *et al.*, 2007). In addition, various datasets obtained from different studies, including hydrological data (i.e. predicting river flow), housing data (i.e. Boston housing prices) and laser data (i.e. time series intensity), were used to compare the performances of three machine learning algorithms: NN, the boosted regression decision tree and RF. The results suggest that no algorithm outperformed others on all datasets (Shrestha and Solomatine, 2006). Furthermore, a comparison between RF and the boosted decision tree showed that the boosted tree algorithm provided better accuracy; however, the RFs appeared to be more stable (Chan, Huang and Defries, 2001). The boosted tree algorithm was found susceptible to being noisy and could consume more time in the training process than the RF (Briem, Benediktsson and Sveinsson, 2002).

Training speed is also of great concern in selecting the appropriate machine learning algorithm to process a large number of datasets. Research has suggested that linear regression, the boosted tree and the RF are considered fast trainers (Rohrer, 2016). A comparative evaluation among NN, RF and SVR in processing facial images was performed, and the time consumed for processing 55,000 pieces of data and cross-validating using the four algorithms was nine hours (SVR), five hours (RF) and nine minutes (NN) (Fernández *et al.*, 2015).

Multi-class classification

IQL-3/4 level road asset management allows for the measured IRI values of road sections to be given in terms of a category (e.g. very good, good, fair, poor, very poor) i.e. classification. Schlotjes (2013) reviewed a number of machine learning classifiers in terms of their advantages and disadvantages. Her review is summarised in **Table 3.6**.

Table 3.6: Summary of machine learning classifiers (Schlotjes, 2013)

Machine learning classifiers	Advantages	Disadvantages
Naïve Bayes	A simple, probabilistic classifier	Assumes strong independence in the model, and as a classifier, this algorithm is reported as a poor performer
Logistic regression	Quick to train with a linear decision boundary that is easily understood, it was found to be better than the decision tree, K-nearest neighbour and the linear model	Weaker algorithm compared to NNs and the SVM.
Neural network	Use nonlinear boundary to solve a complex problem, reported performing well with large datasets and various attributes	Request long training time and can be computationally exhausted; known as a black-box approach. with less interpretability
Support Vector Machine	Found to be more interpretable and requires less training time than NN, it is reported as a good performer with human knowledge that is involved in constructing the model	Less interpretable than the decision tree and logistic regression and slightly considered as a black-box approach
Decision tree	Interpretable and found to be a good performer on small datasets, found to be comparable to NN but better than linear regression	Performs well on easily separable classes but found less successful on complex problems
Random forest	Good performer against NN; improved results on the single decision tree and can handle various data sizes; better	Less interpretability than the decision tree

interpretability than NN and the SVM		
K-nearest neighbour	Simple and user-friendly algorithm and generally better than the linear classification algorithm	Deficient in a complex problem, requires a long time for large datasets and has been found to be computationally exhausted. Poor performer compared to NN and SVM

A comparative study between NNs and RFs was conducted for classifying road conditions using acceleration data obtained by accelerometers mounted in moving vehicles by Jang *et al.*, (2016). A number of variables were selected that considered the vehicle speeds, the accelerometer placement and the peak of acceleration. The results showed that NN performed slightly better than RF (Jang *et al.*, 2016). A comparison of 10 different machine learning algorithms for the binary classification problem was performed, and the results suggested that RF and NN provided the best average performance (Caruana and Niculescu-Mizil, 2006). A comparative study between various machine learning algorithms for road type classification was performed in the UK using driver monitoring sensor data collected on buses, and four road types (Class A, B, C and D) were considered to represent primary roads, small-scale local roadways, unassigned roads and motorways. The performances of different machine learning classifiers (i.e. decision tree, Naïve Bayes, NNs, simple one rule, AdaBoost tree and RF) were compared, and the results suggested that RF performed the best (Taylor *et al.*, 2012). Similarly, research has shown that RF perform better than NN, SVM and Naïve Bayes, with an error of only 6.1% (Kouzani, 2007).

Summary

In general the literature suggests that researchers may find that any machine learning algorithm performs more superiorly to others depending on the specific requirements to hand. The above notwithstanding, the literature suggests that the single decision tree, logistic regression and Naive Bayes approaches in general perform less well than the RF, the boosted decision tree and NN (Caruana and Niculescu-Mizil, 2006).

The above literature review would suggest therefore that a number of machine learning algorithms may be potentially suitable for further consideration in this research. These include:

For the regression problem (determining precise values of road roughness to IQL-2/3):

- Artificial neural network (NNs)
- Random forests for regression (RF)
- Gradient boost regression tree
- Polynomial regression
- Decision tree (DT)
- K- Nearest neighbour (KNN)

The classification problem (classifying road roughness within a range of values; suitable for IQL-4):

- Naïve Bayes

- Random forest (RF)
- Neural network (NN)
- The SVM
- Gradient-boosted tree
- Decision tree
- K-nearest neighbour (KNN) classification

These algorithms are further investigated in **Chapter 6** to select the most appropriate approach that predicts road roughness (IRI) accurately using a smartphone-based application.

Chapter 4

METHODOLOGY

Section 4.1 Introduction

A computerised road management system should consist a of number of components (See *Figure 4.1*), including 1) a comprehensive data collection process which enables essential information (e.g. road inventory, ride quality, pavement strength, traffic loading, etc.) to be obtained; 2) a database where information can be stored a format that can be access by a decision support tool, and; 3) a decision support tool that processes the input data into meaningful management information to support road management decision making (H. Kerali, 2002).

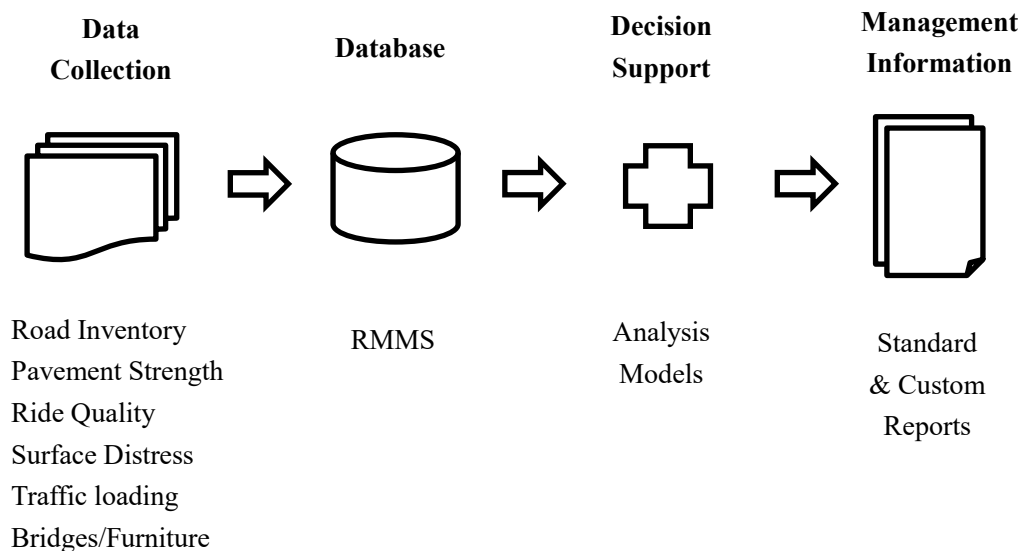


Figure 4.1: Components of a road management system (H. Kerali, 2002)

The goal of this research is to determine if a smartphone can be used to capture road condition data which is suitable for use in such a computerised management system. The research will focus on the types of defects which can be measured, the accuracy to which such defects may be measured, the development of a prototype system for predicting road roughness (and the parameters required by this system in addition to smartphone measurements) and the suitability for the use of such data for decision making at various levels of road management (see [Section 2.3.5](#)). This chapter describes the associated methodology to achieve the above and in relation to the objectives of the research presented in **Chapter 1**.

Section 4.2 Research methodology

The methodology developed in this research is summarised in **Figure 4.2** and described in more detail below:

1. Data requirements for asset management
2. Understanding the smartphone
3. Identifying influencing factors
4. Developing the prototype system

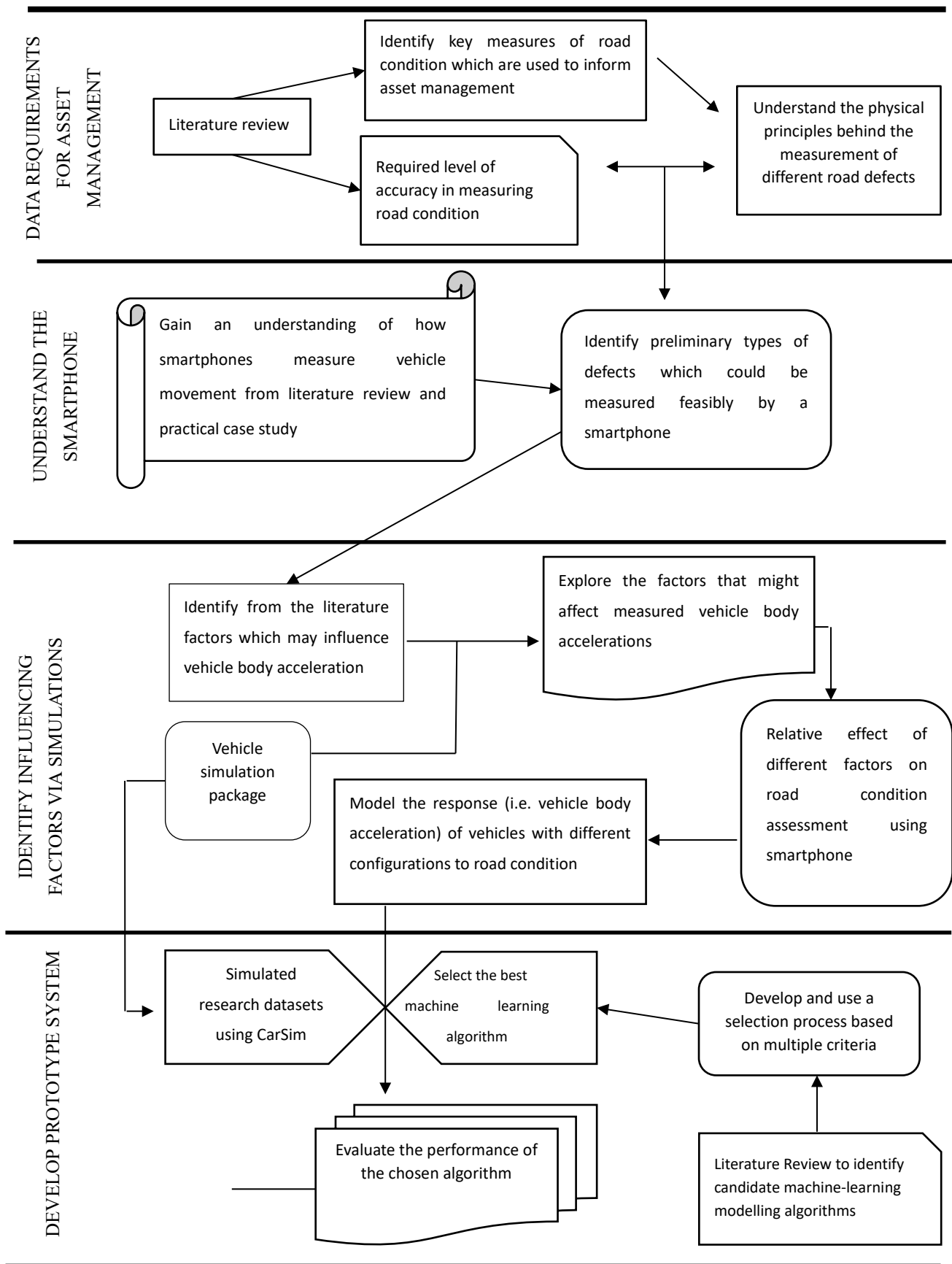


Figure 4.2: Research methodology

Section 4.2.1 Data requirement for asset management

To address objective one, a literature review was used to identify the primary measures of road condition (see [Section 2.3.2](#)), and their associated data requirements, which are needed to make road maintenance decisions. These defects are road roughness, surface distress (e.g. cracking), structural strength, pavement texture and skid resistance. The World Bank Information Quality Level approach (see [Section 2.3.1](#)), was identified as a suitable means of determining the required accuracy of data for particular levels of road asset management.

Section 4.2.2 Understanding the smartphone

To answer the objective two, this section of the methodology explores the possible road condition data that can be measured by using smartphone technology. Firstly the components of a smartphone which allow physical measures to be quantified by the device were explored by means of a literature review. This allied to an understanding of how road defects are recorded using automated means allows for an appreciation of the types of defects which a smartphone measurement system might be able to record. The review of the literature found that road roughness and discrete areas of surface distress (i.e. potholes) could be measured using a smartphone fitted in a moving vehicle (See [Section 2.3.3](#)). This is because the accelerometers fitted inside, a smartphone can measure vehicle vertical body acceleration, and such measured vertical body accelerations are mainly caused by road roughness and defects (e.g. potholes).

In addition, it was also considered useful to gain a practical insight (i.e. how existed smartphone system performs) by undertaking two trials of using the smartphone application to assess road condition. These trials are described in **Chapter 5** and Sections 6.2.3 respectively.

Conceivably it might also be possible to assess the skid resistance and structural strength/residual life of roads using smartphone based technology with careful processing of the data. Although the measurement of these defects are not addressed in the experimental part of this research, further consideration is provided in the discussion part of the thesis (see Section 9.3).

The reason of focusing on roughness is because it is a particularly useful measure of road condition as it is considered as the most important indicator of ride quality and therefore road use cost and therefore it is used as the basis of a number of road investment appraisal tools (e.g. HDM-4, see **Appendix B**). Road roughness can also provide useful information in helping to diagnose surface distress and structural deterioration, and its measurement can be automated (Robinson *et al.*, 1998).

Section 4.2.3 Identifying Influencing Factors

As described in Section 2.3.3 most smartphones have inbuilt accelerometers which allow the measurement of acceleration in three dimensions. The measurement of the vertical acceleration potentially allows the vertical profile of a road to be determined via a smartphone fitted inside a moving vehicle. Section 2.3.4 discusses how such

measurements can be translated into IRI. However, the literature showed that there are a number of factors which can influence the measurement of the vertical acceleration of a smartphone when fitted in a moving vehicle. These include the smartphone itself, vehicle related factors and the shape of the road profile.

In order to investigate these factors it was decided to use an experimental approach based on a number of numerical simulations carried out using a vehicle dynamics package (see below). Such an approach was chosen due to the wide variety of factors involved and therefore the number of experiments required. It was felt that carrying out the potentially large number of numerical simulations would be much less time consuming than approach which tried to obtain data from actual smartphones fitted inside a moving vehicle and that it would be easier to control for all of the independent variables. A disadvantages of this approach is that actual smartphone data is not used. Also it is implicitly assumed that the direct output of the vehicle simulation package, namely vertical vehicle body acceleration, is equivalent to the vertical acceleration measured by the accelerometer in a smartphone. This is because data gathered from smartphone fitting in a vehicle can be used as a proxy for those from car-mounted sensors (e.g. accelerometer) (Qiu *et al.*, 2018; Paefgen *et al.*, 2012), while CarSimTM has been validated using car-mounted sensor data measured in real vehicle (Mechanic Simulation, 2017). This work is described more fully in **Chapter 6**, whilst choice and application of the vehicle dynamics package is described in the section (vehicle-related factors) below.

Smartphone associated factors

The smartphone associated factors identified by the literature review (**Chapter 2**) include the characteristics of smartphones (e.g. operating range, data sampling interval), smartphone measurement errors, position inside the vehicle and effect of the smartphone type on measuring road roughness.

Vehicle-related factors and road profile

In order to study the vehicle related factors which could influence the measurement of vehicle body vertical acceleration a suite of simulations of different types of vehicles traversing a number of road profiles will be carried out using a vehicle simulation software known as CarsimTM. CarSimTM is a 3-dimensional non-linear vehicle dynamic modelling software that is widely used commercially as the basis of laboratory testing and vehicle dynamic simulation (Mechanic Simulation, 2017).

CarsimTM was chosen for the task in hand for the following reasons:

1. The software is a preferred tool by the industry for modelling vehicle dynamics (Dupuy *et al.*, 2001; Benekohal and Treiterer, 1988).
2. The software has been shown in the literature to predict accurately the vehicle dynamic performance of a large number of standard vehicle types.
3. It can simulate a number of actual vehicle classes whose physical parameters (e.g., suspension stiffness, sprung mass and tyre pressure, etc.) and operating speed can be varied.

4. Any road profile can be modelled in the software
5. The outputs provided are ideal for the needs of this research, and easily available.

The outputs include vehicle sprung mass accelerations, displacement, suspension force and steering force.
6. The data provided in CarSim™ has been validated using sensor data measured in real vehicles.
7. Data can be exported from CarSim™ for further analysis in a format which is easily analysed by other computer packages.
8. The software is available within the School of Engineering at the University of Birmingham

A parametric study was carried out using CarSim™ vehicle simulation to know the relative effects of factors that are associated with vehicle (See [Section 6.3](#)) and road profile (See [Section 6.4](#)) on the vehicle body acceleration measured. This was achieved by changing each factor (within a range of values) in turn, and keeping all other factors constant. As mentioned above, the vehicle sprung mass body acceleration of the simulated vehicle response simulated on known road profiles using CarSim were collected for the purpose of the analysis. Following the review of the literature (see [Section 6.3](#)), the vehicle-related factors chosen for investigation were.

- Vehicle speed
- Vehicle type

- Suspension stiffness and damping ratio
- Tyre pressure
- Sprung mass
- Driving style
- Wheelbase

For each simulation the Root-Mean-Square of the measured vehicle body accelerations (Grms) obtained from CarSimTM was compared with the actual value of road roughness of the road profile used for the simulation. The Grms was used for this purpose because research has shown that the power spectral density (PSD) of a vehicle's vertical acceleration, obtained from frequency domain analysis, correlates closely with road roughness (Hesami and McManus, 2009; Sun, 2001; Marcondes *et al.*, 1992). According to Parseval's Theorem the Root-Mean-Square (Grms) of the vehicle body vertical acceleration in the time domain is equal to the square root of the integral of the PSD (See Section 2.3.4) (Van Baren, 2012; Rogers *et al.*, 1997). The second reason for choosing Grms is that the Grms is calculated over a time series acceleration data at a fixed interval (every 0.1 mile) and this can reduce the dimension of data storage and difficulties in processing large dataset.

The road profiles for input into CarSimTM were acquired from actual road sections provided in the LTPP dataset (FHWA, 2018). The LTPP dataset was established and is managed by the U.S. Federal Highway Administration. Data of approximately 2, 100 road

section across the North America were obtained. The road profiles were chosen to represent road conditions ranging from very good (i.e. IRI = 0 to 2.5 m/km) to poor (i.e. IRI = 6 to 10 m/km). The IRI of each section was determined from the profile of the section using a computer program written by the researcher in Visual Basic for Applications (VBA).

Modelling the response

To better understand the impact of the above influencing factors on the measurement of road roughness mathematical models were built to predict IRI from vertical vehicle body acceleration (Grms) data together with information associated with the influencing factors. Two modelling approaches were trialled, one was based on a conventional multivariate linear regression and the second was a machine learning approach. One algorithm was built using the former approach and two using the latter. This is further described in **Chapter 6**.

Section 4.2.4 Developing the Prototype System

Objective four of this research is associated with the development of a prototype system that can utilize representative smartphone data to determine road condition (IRI). To this end the approach chosen was to identify and trial a suitable algorithm which can process the acceleration data received from a smartphone to determine IRI, possibly utilizing knowledge of the factors influencing the measurement of acceleration (e.g. vehicle speed).

Chapter 3 suggested that suitable category of algorithms which might be appropriate for

this task are those known as machine learning algorithms. In order to choose an appropriate machine learning algorithm for the task at hand, a number of candidate algorithms were chosen from the literature and the most appropriate was selected based multiple criteria. Five weighted criteria namely: performance, training speed, suitability for large datasets, the ability of generalisation and interpretability were chosen to assess the overall performance of each algorithm. The weightings of the criteria were determined based on expert knowledge and the review of corresponding algorithm in literature. This process is described in more detail in Chapter six.

For the development of the machine learning algorithms the KNIME software was used (KNIME, 2018). KNIME has a number of standard machine learning algorithms and was chosen for this task for the following reasons:

1. KNIME is an open-source software and therefore free to use
2. KNIME has a graphical interface software which is easy to use and it is also compatible widely available software which can be used for statistical analysis including Microsoft ExcelTM and WekaTM

Data

A dataset of 807, 344 rows was obtained using CarSimTM to train and test the prototype system (see **Chapter 8**). The evaluation of the system was divided into a regression problem to try to determine IRI as exactly as possible and a classification problem by

which IRI was predicted in terms of ranges of values. For the regression problem the associated information quality level is IQL-2 i.e. that required for project level or detailed programme management (See Section 2.3.1). While for the classification problem the relevant information quality level is IQL-3/4, the predicted IRI is categorical (e.g. good, fair and poor) i.e. that considered suitable for network-level strategic analysis.

Discuss the Usefulness of Data Provided by the Prototype System

A wide discussion is included of the lesson learned in the development of the prototype system (see **Chapter 9**), as well as about the wider application on how the data obtained from the prototype system may be used to support road asset management (cf. **Appendix B**). A road management system which can be used to identify and prioritise candidate road sections for treatment should include: a) the determination of appropriate maintenance standards and associated maintenance expenditure. These can be considered to be tasks associated with strategic planning at the network-level. This process requires data at IQL-3/4. b) The resulting maintenance standards can then be input in a computerised Road Maintenance Management System (RMMS), along with road condition data measured by smartphones on a recurrent basis (e.g. annually). The comparison of actual measured condition with the standards allows a list of defective road sections requiring treatment to be produced. This process can be regarded as project to network level management. (Odoki and Kerali, 2000).

Section 4.3 Summary

This chapter has described the research methodology developed to build a prototype system which can determine road roughness from data collected from smartphones fitted into a variety of vehicles. The methodology consisted of:

1. Desk-based research to understand the primary measures of road condition used for asset management decision making, and their associated required information quality levels (See Section 2.3.3).
2. The selection of road roughness as the focus of the research
3. Understanding of how the accelerometers built into many smartphones may be used to measure acceleration and therefore road roughness from a review of the literature and by a number of practical experiments.
4. A literature review of possible influencing factors which may affect the measurement of vehicle body acceleration using smartphone-based applications. Further analysis of the influencing factors via a parametric study carried out with the aid of the vehicle dynamic simulation package CarSim™.
5. The selection of computer algorithms to develop predictive models which can compute road roughness (IRI) from the measured vehicle body accelerations of fleets of vehicles. Emphasis was given on the selection and development of machine learning algorithms which appear to be suitable for network-level data measurement and can be used on datasets which contain, other than measured Grms, a minimal amount of information.
6. Testing the prototype system using artificial datasets obtained from CarSim™

simulations (cf. Section 8.3) to find out the level of accuracy such system can achieve.

7. Discuss the usefulness of the output from the system (IRI) in asset management (see

Chapter 8 & 9).

Chapter 5

CASE STUDY: INVESTIGATING THE ACCURACY OF MEASURING ROAD ROUGHNESS USING AN EXSITING SMARTPHONE-BASED SYSTEM

Section 5.1 Introduction

A review of the literature has summarized a number of smartphone-based applications (see [Section 2.3.6](#)). To gain a practical understanding of collecting road roughness data using a smartphone, the condition of a 12.8 km section of the UK road network (A25 in Kent) was measured using a widely available smartphone application (Roadroid) (Forslöf, 2012) (see *Table 2.5*). Although the intention was not to carry out a controlled experiment to determine the accuracy of Roadroid, but rather to gain an better understanding of the practicalities of using a smartphone, the results obtained from Roadroid were compared with historical data obtained using a laser-based SCANNER (i.e. ‘Surface Condition Assessment for the National Network of Roads’) measuring system (UK Roads Board, 2009). SCANNER is the UK’s standard procedure for measuring road condition on its strategic road network. It does not measure IRI, but rather the longitudinal variance (LPV) of the road profile. Nonetheless, it was hoped that trends in road condition measured by the two approaches could be compared.

Section 5.2 Methodology

The methodology adopted compared road roughness values obtained using Roadroid with those values determined using the UK Pavement Management System (UKPMS) standard procedure for the measurement of road deterioration (i.e. SCANNER) (Halcrow Group Limited, 2011). For this purpose, data were collected using a Smartphone fitted with Roadroid on the A20 on 5th January 2016. The Smartphone was mounted in a light van in transit, and the same route was traversed three times (See *Figure 5.1*).





Figure 5.1: Experiment undertaken using Roadroid, A25 in Kent, UK, 2016

Section 5.2.1 Roadroid Application

A commercial Smartphone application known as Roadroid claims to be able to measure road roughness accurately. In principal, a Smartphone mounted inside a moving vehicle can measure the vertical acceleration of a vehicle body via the accelerometer within the Smartphone (cf. [Section 2.3.3](#)). Roadroid processes the vertical acceleration data using algorithms hidden from the user, but which require user calibration according to the type of vehicle and smartphone. The algorithm is also claimed to be able to cope with the effects of vehicles moving at different speeds.

To obtain road roughness data using the Roadroid system, the Roadroid app was running on a Samsung Galaxy S3 device that was installed on the dashboard of a light survey van (see

Figure 5.1). The Roadroid app was used with the 4WP Jeep and Smartphone type 9100S settings and a calculation sensitivity of the default value.

Accordingly, three runs of road roughness measurement using the smartphone app were undertaken on the 12, 840 m length of road on 5th January 2016. The average vehicle speeds of three runs were 62.04, 61.73 and 62.97 km/h respectively.

Section 5.2.2 SCANNER Data

The SCANNER method collects a number of parameters associated with road condition—one of which is to do with the variance of the longitudinal profile (LPV). The LPV measures how much the road undulates along the direction of travel and is considered to be a good indicator of road roughness. For further details, see *SCANNER Surveys for Local Roads* (Halcrow Group Limited, 2011).

In this study, the SCANNER 3m enhanced LPV was selected on both the left (LL03) and right (LR03) sides to represent the road longitudinal profile data. The SCANNER data were obtained from a survey undertaken by Highway Surveyors in June 2014.

Data collected from the left and right wheel tracks were averaged to obtain a single value to be compared with the values obtained from Roadroid. 3m LPV data are the three-metre moving average of the road profile, and therefore are not an exact representation of IRI. However, a number of empirical equations have been developed to convert 3m LPV to IRI. The correlation shown in *Equation 5.1* has been used herein to convert 3m LPV values to their equivalent IRI values (Bradbury *et al.*, 2012).

$$LPV(3m) = 0.2117 IRI^{1.8507}$$

Equation 5.1

Where: LPV (3 m) is the 3 m longitudinal profile variance (mm^2) and IRI is the International Roughness Index (m/km).

Section 5.2.3 Experiment Locations

A 12,840 m section of the A20 in Kent, UK was selected to compare the measured outputs from Roadroid and SCANNER (See **Figure 5.2**). The measurement started at the Geographic Information System (x, y) co-ordinate of 553008.5, 167519.3 and ended at the point (x, y) of 561636.9, 159153.5. (See the blue marks in **Figure 5.3** and **Figure 5.4**).

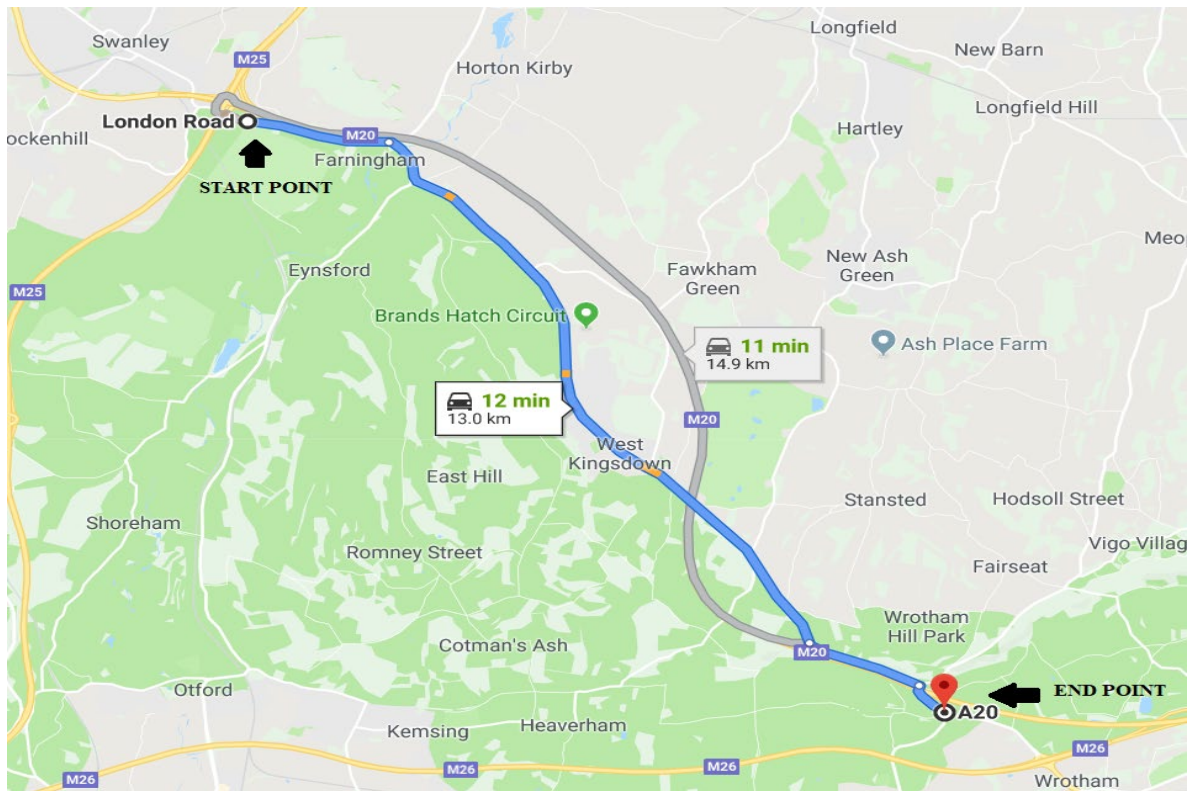


Figure 5.2: Entire route of the experiment on A20, Kent, UK



Figure 5.3: Starting point of the experiment on A20, Kent, UK



Figure 5.4: Starting point of the experiment on A20, Kent, UK

Section 5.3 Performance of Roadroid Application

The following approaches were used for comparing the performance between Roadroid and the SCANNER system. The numerical IRI values measured from these two systems were compared:

1. Visual observation
2. Statistical Analysis
3. Threshold Analysis
4. Frequency domain spectral analysis

The comparisons described above will focus on the entire road length measured as a whole, whereas detailed analysis on each section can be found in **Appendix C**.

Section 5.3.1 Visual Observation

The IRI values obtained using Roadroid and SCANNER were shown in **Figure 5.5–5.7** for visual observation. For the Roadroid values, Run One (R1), Run Two (R2) and Run Three (R3) represented each of the three repeated measurements along the same route.

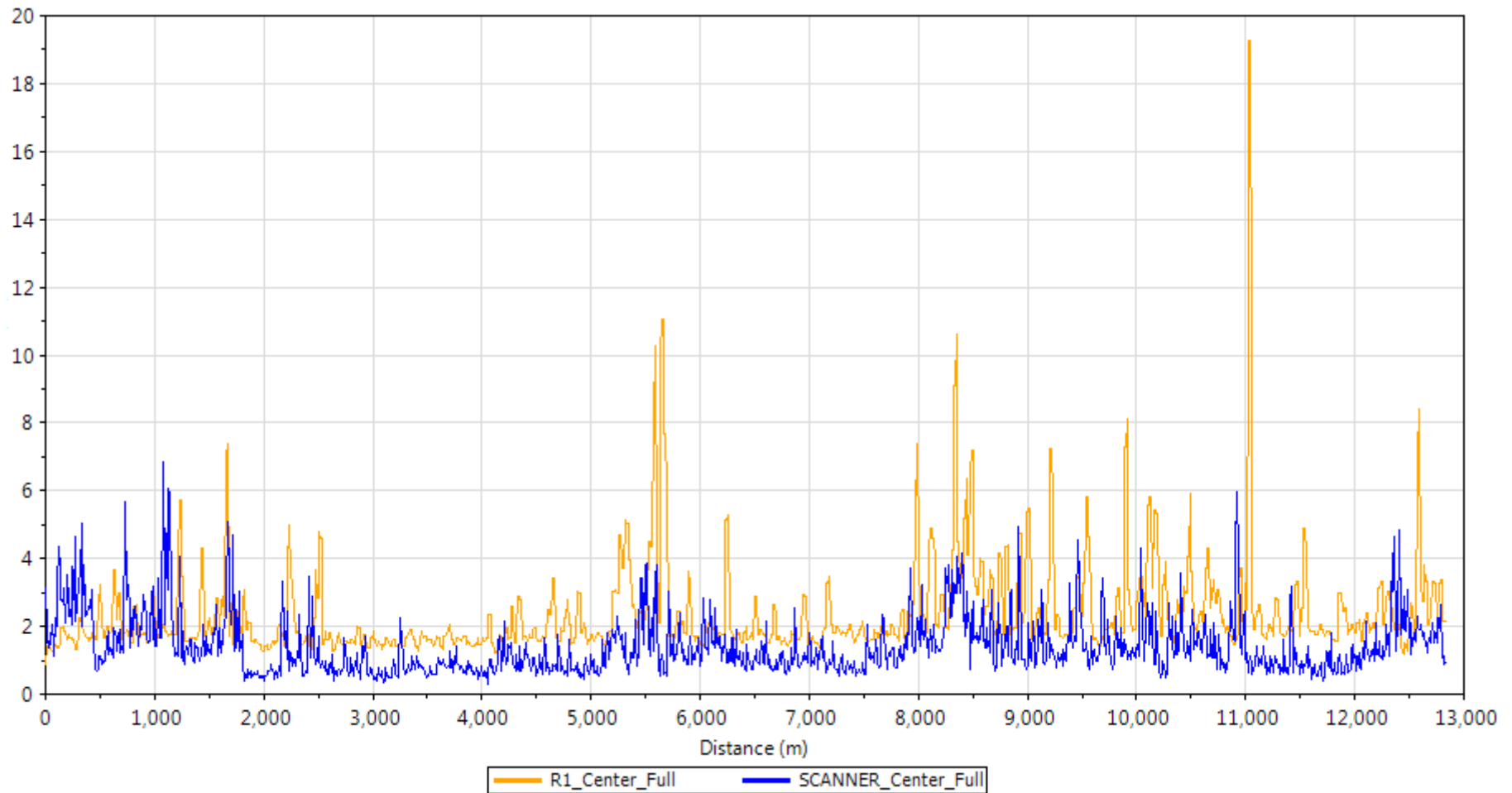


Figure 5.5: Comparison of road roughness data measured by SCANNER and those measured by Roadroid (Run One)

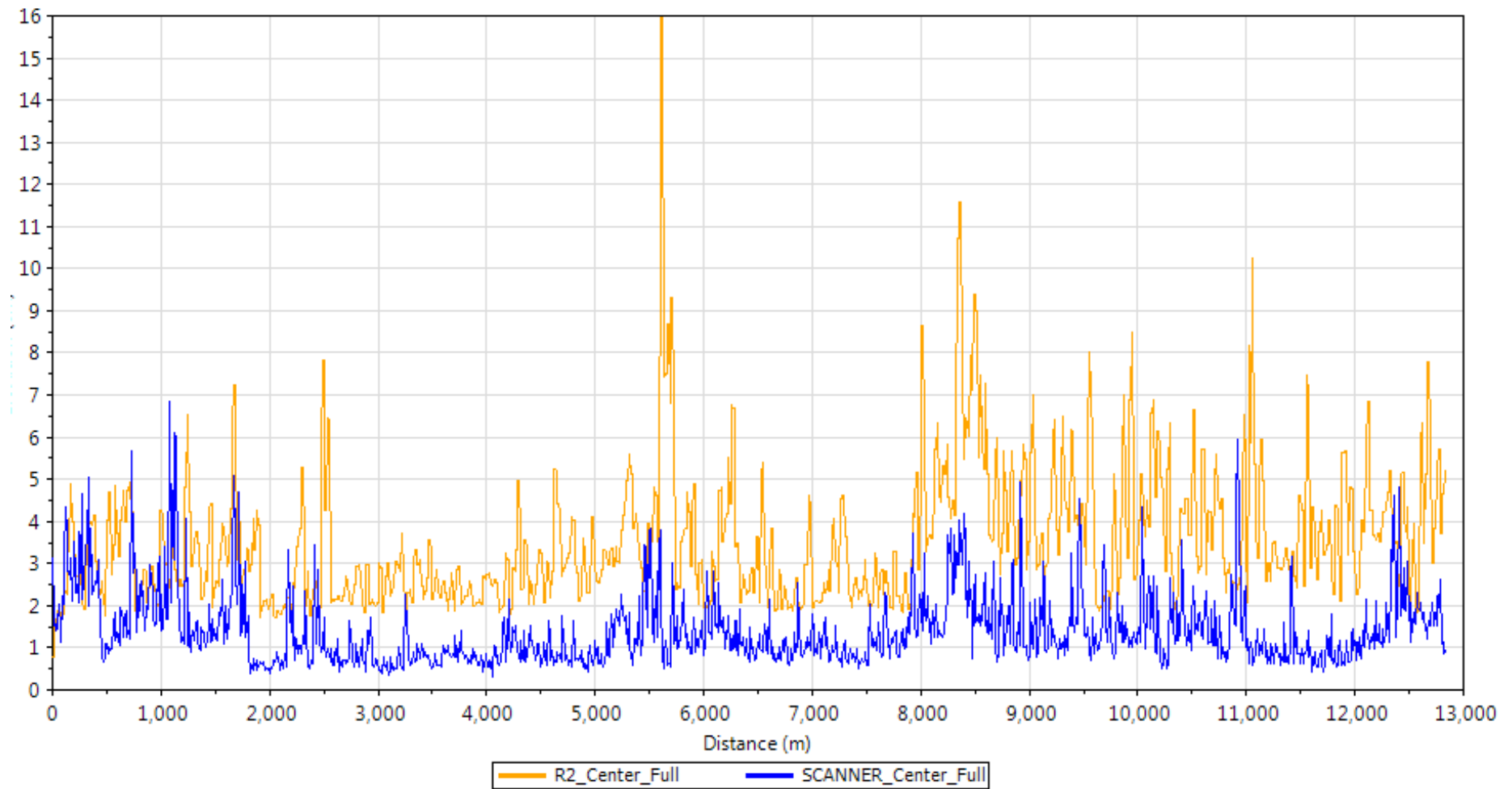


Figure 5.6: Comparison of road roughness data measured by SCANNER and those measured by Roadroid (Run Two)

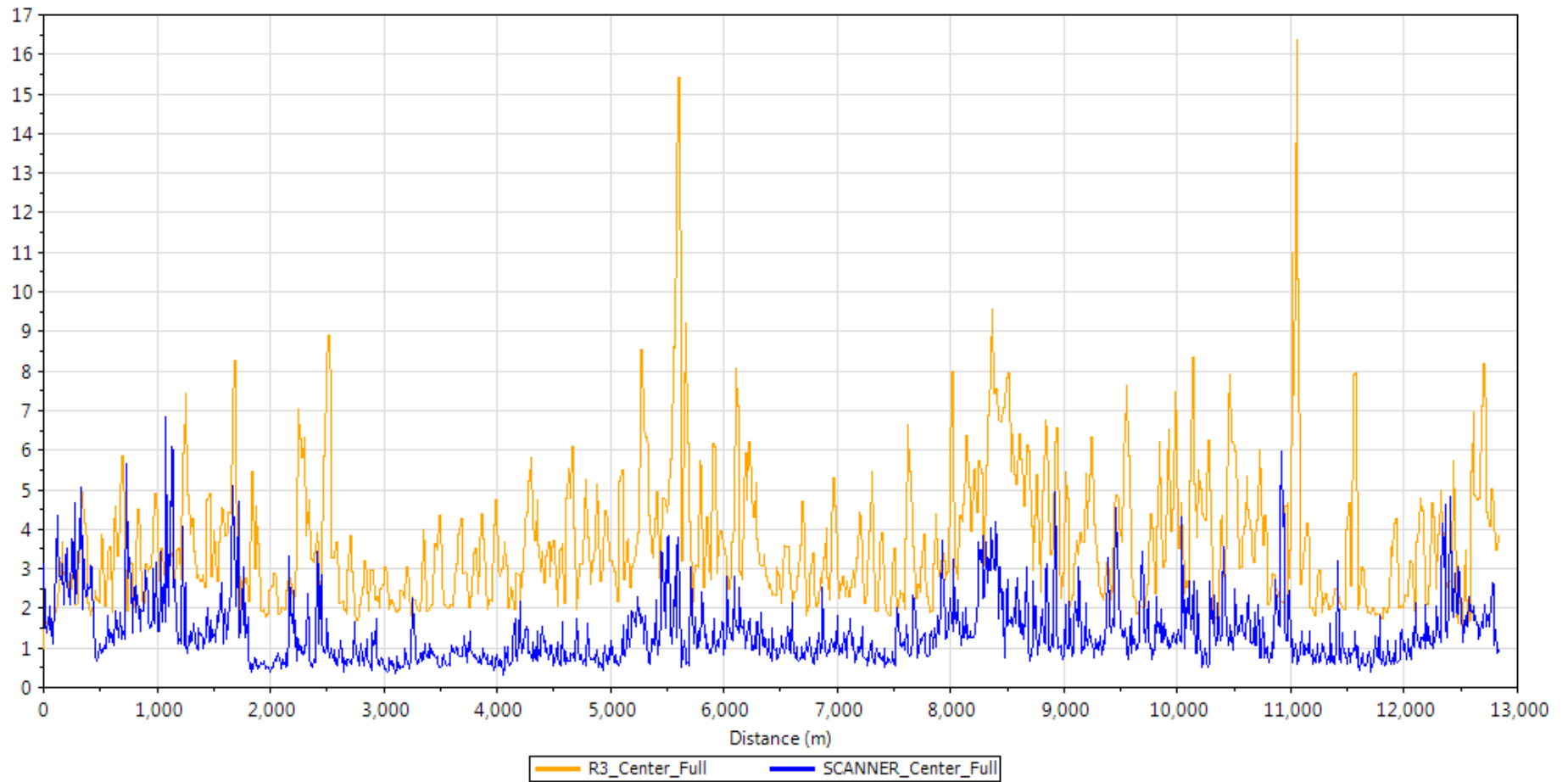


Figure 5.7: Comparison of road roughness data measured by SCANNER and those measured by Roadroid (Run Three)

As can be seen from **Figures 5.5–5.7** there appears to be some similarity between the locations of the peaks of the Roadroid and SCANNER data sets — i.e. peaks and troughs appear to correspond (e.g, rough road section before Chainage 2,000m, near 6,000m, from 8,000m to 10,000m, around 11,000m). However, the IRI values measured by Roadroid were generally higher than those measured by SCANNER.

It is also apparent that the results for R1 appear to be inconsistent with those from R2 and R3. In fact, the data gathered from R1 correlate better with SCANNER data. This observation may be due to the fact that Roadroid R1 was the first trial of the experiment and that perhaps the driver was more restrictive on the speed control of the vehicle and the vehicle was operating at more consistent speed.

Section 5.3.2 Statistical Analysis

Table 5.1 presented a summary of some basic statistics related to the SCANNER and Roadroad data sets.

Table 5.1: Statistical analysis of data gathered from Roadroid and SCANNER

	SCANNER	Roadroid Application		
		R1	R2	R3
<i>Average of measured IRI (m/km)</i>	1.396	2.40	3.55	3.67
<i>Variance (m/km)</i>	-	1.00	2.15	2.28
<i>Ratio of Error</i>	1	2.14	3.19	3.30

Where:

Average of measured IRI = Average road roughness measured over the entire road section.

Variance = Average of (IRI measured from Roadroid minus the calculated IRI from SCANNER data)

Ratio of Error = Average of (IRI measured from Roadroid divided by the calculated IRI from SCANNER data).

As can be seen from **Table 5.1**, the Roadroid data set for R1 has the closest average measured IRI value when compared with the SCANNER data (albeit 72% higher on average). R1 also has the smallest *Variance* and *Ratio of Error*. R2 and R3 were found to give average IRI readings that were 155% and 169% higher than those determined using SCANNER.

The Ratio of Error shows that the average ratios of Roadroid data sets to the SCANNER data set were 2.14, 3.19, and 3.30 respectively. A further analysis was carried out to compare the ratio of individual measurements. By comparing each data point, for the ratio of R1 to SCANNER, it was found that only 19% of the ratios were between 0.7 and 1.3 (i.e., within 30%) and only 13% were between 0.8 and 1.2 (i.e., within 20%).

To determine whether the differences between the SCANNER and Roadroid measurements were significant, a paired sample t-test was carried out between SCANNER and Roadroid R1 datasets. The results are shown in **Figure 5.8**.

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 IRI	1.3964	1284	.85190	.02377
R1	2.3955	1284	1.47736	.04123

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 IRI & R1	1284	.131	.000

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 IRI - R1	-.99919	1.60588	.04482	-1.08711	-.91126	-22.295	1283	.000

Figure 5.8: Paired sample t-test results of SCANNER data against Roadroid R1 data.

This analysis demonstrates that the difference between the two sets of readings is statistically significant ($\text{sig} < 0.05$)—i.e., not due to chance—and it is of statistical importance (to the 95% confidence level) that the Roadroid readings are on average between 0.91126–1.08717 m/km higher than the SCANNER data.

The apparent anomalies between the magnitudes of the readings may be due to the following factors:

1. Data gathered by SCANNER are from the year 2014, while data collected by the Roadroid app are from January 2016. During the intervening two-year period, the road will have deteriorated and therefore it can be expected that the Roadroid data, on average, will show a higher IRI value. This proposal was confirmed via the t-test described in **Figure 5.8**. However, the *t-test* suggested that on average the

difference of IRI between the two data sets is 1 m/km which the literature would suggest is a large amount for a road to deteriorate in two years.

2. The use of *Equation 5.1*, which is based on empiricism.
3. Roadroid has a feature that allows calibration to different classes of vehicle and smartphone types. It may be possible, therefore, that an alternative calibration to that used may have given improved results (in terms of the differences between the magnitudes). Further, there is an additional feature on Roadroid that allows the sensitivity of the measurement to be adjusted. For this research, the default value (i.e., 1.5) was used, and an alternative value may have improved the match between the Roadroid and SCANNER data sets.

Statistical Comparison of Data Measured by Roadroid

Figure 5.9 presents a statistical comparison of the three Roadroid data sets. The three sets of data have average IRI values of $R1 = 2.39$ m/km, $R2 = 3.55$ m/km and $R3 = 3.67$ m/km respectively. There is a degree of correlation between the three data sets (paired correlation values of 0.472, 0.657 and 0.552 respectively). $R2$ and $R3$ have the highest paired correlation, which corroborates the similar average IRI values of the $R2$ and $R3$ data sets (the mean difference between the two data sets is only 0.12m/km). This result suggests a level of repeatability using this system.

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	R1	2.3955	1284	1.47736	.04123
	R2	3.5502	1284	1.62083	.04523
Pair 2	R2	3.5502	1284	1.62083	.04523
	R3	3.6733	1284	1.72535	.04815
Pair 3	R1	2.3955	1284	1.47736	.04123
	R3	3.6733	1284	1.72535	.04815

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	R1 & R2	1284	.472	.000
Pair 2	R2 & R3	1284	.657	.000
Pair 3	R1 & R3	1284	.552	.000

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	R1 - R2	-1.15461	1.59627	.04455	-1.24200	-1.06722	-25.919	1283	.000
Pair 2	R2 - R3	-.12314	1.38845	.03875	-.19915	-.04712	-3.178	1283	.002
Pair 3	R1 - R3	-1.27775	1.53197	.04275	-1.36162	-1.19388	-29.887	1283	.000

Figure 5.9: Statistical comparison among R1, R2 and R3 datasets

Section 5.3.3 Threshold Analysis

To gain a better idea of the match between the SCANNER and Roadroid data sets (as mentioned in [Section 5.3.1](#)), a comparison was made of the two data sets where the data were found to have exceeded 2 m/km, 3 m/km, and 5 m/km. The purpose of this procedure was to better understand to what extent the IRI curves measured by the two systems were matched in terms of the locations of road sections with high recorded IRI values.

1. Threshold of 2m/km (R1 vs SCANNER):

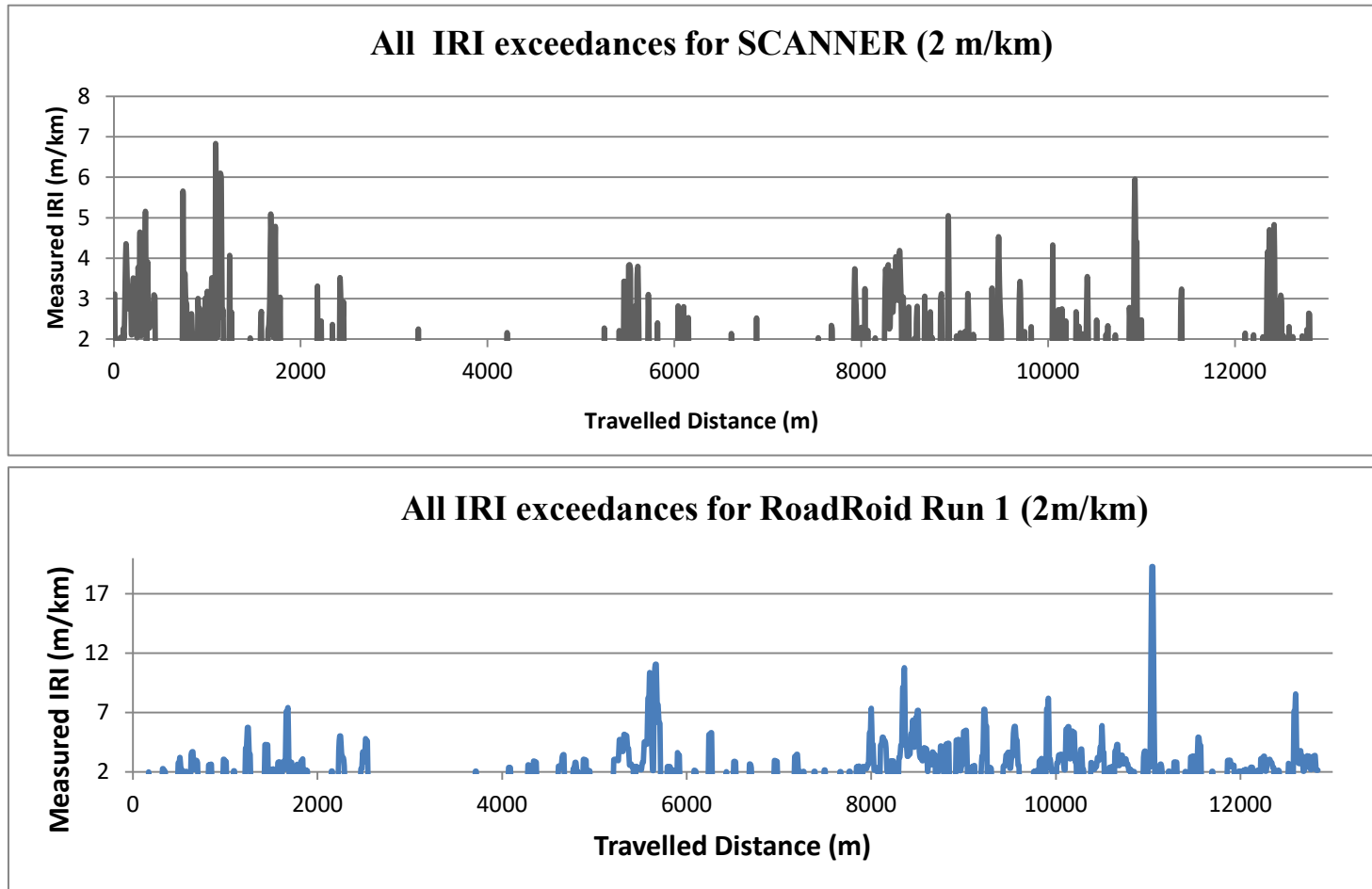


Figure 5.10: Comparison of 2m/km threshold analysis (R1 and SCANNER)

Figure 5.10 presents a comparison of the two data sets, in which it may be seen that there appears to be some similarity in the position of the peaks. Whilst the magnitude of the calculated IRI values appear to be similar for the first 2,000 m, around 5,000 m, between 8,000–10,000 m and after 12,000 m the values determined using Roadroid are significantly higher. These observations are also evident when the threshold is set at 3m/km as shown in **Figure 5.11**.

As can be seen from both **Figure 5.10** and **Figure 5.11**, it is evident that it is possible for Roadroid to detect the changes in road roughness, and indeed nearly all of the individual sections of road with IRI values higher than 3 m/km (as measured by SCANNER) can be detected using the Roadroid system. However, phase differences still existed between the locations of measurements gathered by the two systems due to error in GPS positioning. This phase difference is also shown in **Table 5.2** where the locations and values of the maximum IRIs measured by the two systems were compared. Generally, the locations of maximum IRI values for both systems of measurements were not the same.

The IRI values measured by Roadroid are nonetheless significantly higher than the values measured by SCANNER, a result which could indicate that Roadroid may over-estimate the IRI on rough road sections (over- sensitive).

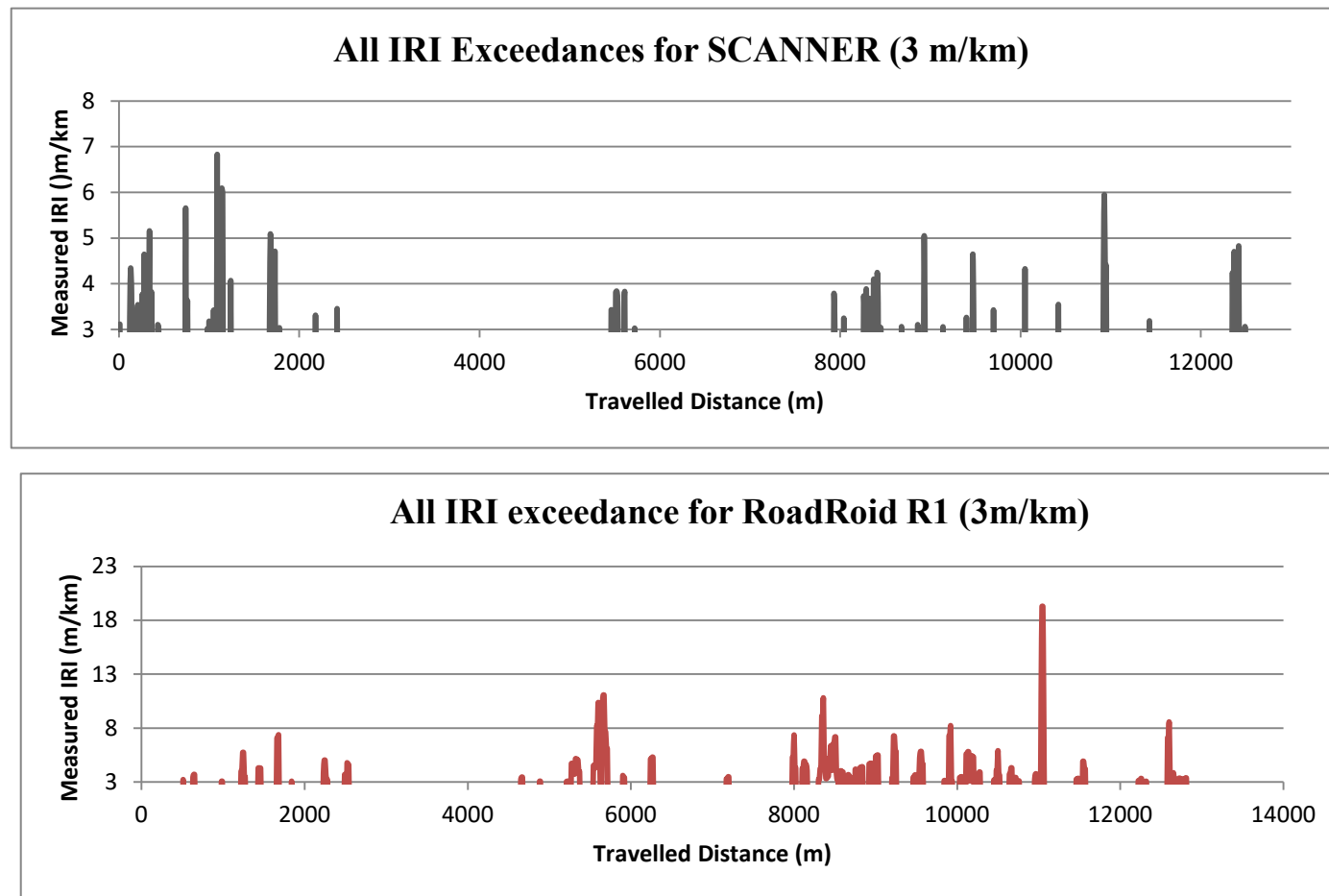


Figure 5.11: Comparison of 3 m/km threshold analysis (R1 and SCANNER)

Table 5.2: Summary of the locations of peak values of IRI measured by two systems

Chainage (m)	R1		SCANNER	
	Peak	Locations	Peak values	Locations
	values	(m)	(m/km)	(m)
	(m/km)			
0 to 3000	7.38	1680	6.83	1090
4000 to 7000	10.28	5600	3.84	5520
7000 to 10,000	10.61	8360	4.92	8930
10,000 to 13,000	19.28	11040	5.95	10930

Similar findings were also found when comparing the Roadroid R2 and R3 data sets with SCANNER data.

Section 5.3.4 Spectral Analysis

An analysis in the frequency domain, in the form of power spectral density (PSD), was undertaken as described below. There were two views from which to compare the PSD curves of the Roadroid and SCANNER data.

The first is the amplitude of the signal for different frequencies, or the area under the PSD.

That is, the higher the area (higher amplitude), the higher the energy (i.e., power) of the

signal (and the higher the amplitude and, therefore, the higher the input (IRI) stimulus). The second view is to compare the shape of the PSD curves under different frequencies because, ideally, if two signals are identical or particularly similar in the time domain, they will have exactly the same—or parallel—PSD curves.

In the first situation, for comparison, the root mean square (RMS) value of signal in the time domain was used to represent the area under the PSD curve. According to Parseval's theorem, they are equivalent measures of signal energy (cf. [Section 2.3.4](#)). The results have been shown in **Table 5.3**, where it can be seen that the Roadroid data exhibit higher energy than the SCANNER data. This result is corroborated by the PSD curves plotted in **Figure 5.12 to 5.14**, which show that the Roadroid data generally have a higher amplitude than the SCANNER data for all frequencies.

Table 5.3: Root mean square value of measured IRI values from Roadroid and SCANNER

	<i>RMS of IRI (m/km)</i>
SCANNER	1.636
<i>Roadroid R1</i>	2.814
<i>Roadroid R2</i>	3.902
<i>Roadroid R3</i>	4.058

In addition, by inspecting the PSD curves shown in **Figure 5.12–5.14**, it was found that the PSD curve for R2 appears to most closely match the PSD curve for the SCANNER data.

This observation also indicates that the shape of the IRI curves for R2 were also the most similar to those of SCANNER in the time domain. Also, it can be seen that the PSD curve for Roadroid has higher amplitude in the low wavelength (<500 m/cycle) and higher amplitude in the high wavelength (>3000 m/cycle). In the time domain, Roadroid has higher values for the ‘short waves’ and ‘long waves’ in the measured IRI curves compared s to those measured by SCANNER.

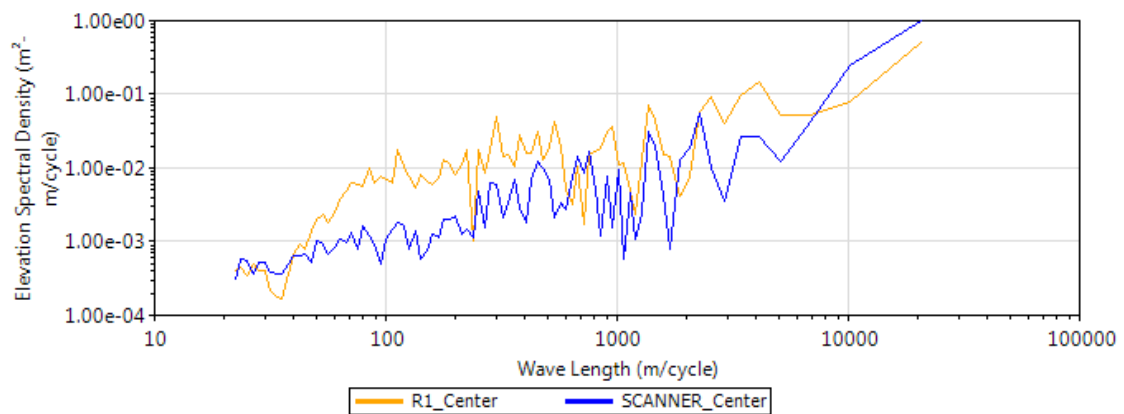


Figure 5.12: Spectral density analysis of Roadroid R1 and SCANNER

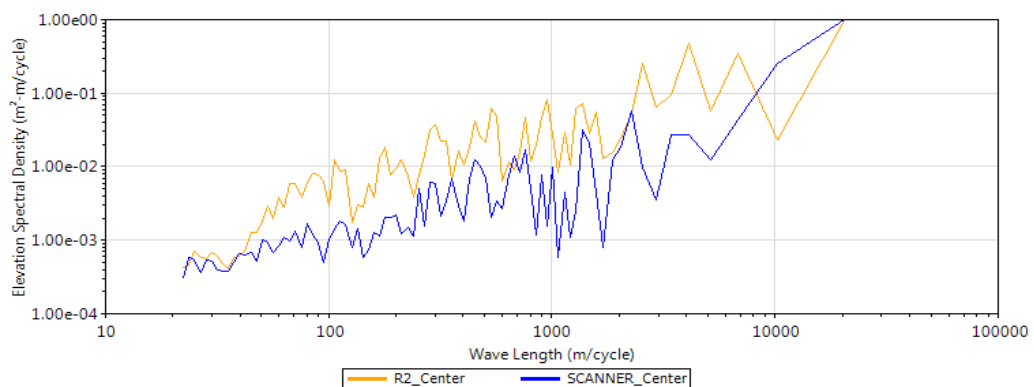


Figure 5.13: Spectral density analysis of Roadroid R2 and SCANNER

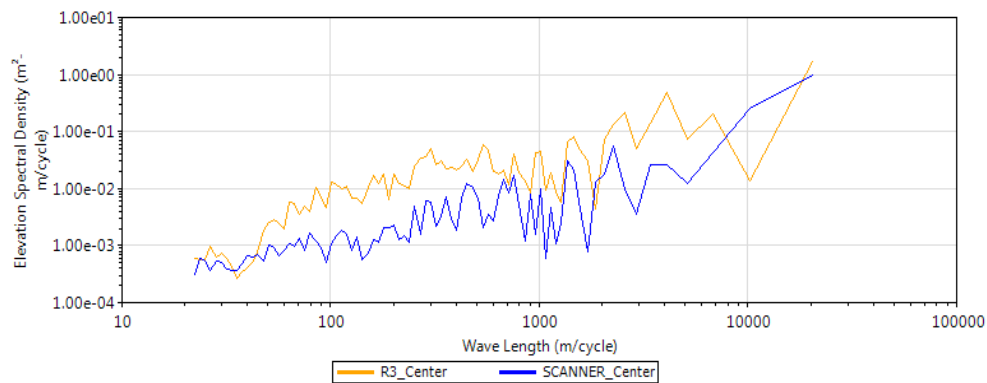


Figure 5.14: Spectral density analysis of Roadroid R3 and SCANNER

Section 5.4 Summary

Despite the limitations of the experiment, it was found that the smartphone system can detect changes in roughness along the road, but the accuracy of the system has not been demonstrated. The data processing algorithm used within Roadroid is based on multilinear regression, which this research found may not be the most appropriate computational tool for the assessment of road roughness from data obtained from a smartphone based system.

A number of findings can be drawn from the above analysis:

1. The Roadroid and SCANNER data do not match perfectly.
2. In general, it would appear that the Roadroid smartphone system can detect changes in roughness, but the accuracy of this measurement is unknown (by the analysis presented here)
3. The IRI values determined using the Roadroid system were generally higher than the SCANNER data in this study

4. The data sets of R2 and R3 using Roadroid generally correlated well with each other (suggesting that the system is repeatable).
5. There are a significantly large number of differences between the measured IRI data from Roadroid and SCANNER, suggesting that further investigation regarding the suitability of the Roadroad system is required.

To discuss the results, a number of aspects need to be mentioned:

1. The data collected by SCANNER were obtained in 2014 and the data measured by Roadroid are from 2016. There is, therefore, a probability that the road section will have deteriorated during the intervening years, leading to higher IRI values being measured by Roadroid than by SCANNER. The results of the analysis tend to corroborate this supposition. There is also a possibility that discrete defects—such as potholes—that were present in 2014 may have been repaired by 2016.
2. Ideally, the two methods of data collection should start with the same location and end at the same point as well; however, there were still some phase differences in terms of the measured chainage, potentially also influencing the similarity of the shapes of the measured IRI curves.
3. The Roadroid app appears to be highly sensitive to the vehicle-type setting selected, the speed driven and the smartphone type selected. These factors need to be carefully considered before the Roadroid system is used in practice. However, it is recognised that these settings could be adjusted to obtain an improved IRI value.

4. The equation selected to convert the 3 m LPV to IRI was based on an empirical correlation and it is recognised that it may not have been especially accurate for the particular conditions at hand.

The findings of the above analysis are helpful in better understanding how existing smartphone based systems may perform in practice, addressing ‘objective two’ in this thesis: that smartphone systems have the potential to measure road roughness. Furthermore, analysis from this chapter has also pointed out that data modelling systems based on multilinear regression can be particularly sensitive to vehicle type and speed, leading to poor stability of the system. This issue has led to fundamental research being carried out systematically, using a vehicle simulation package to understand the relative importance of various influencing factors that affect the measurement of vehicle body acceleration (see following chapter). Additionally, a machine-learning algorithm is introduced for data modelling, as opposed to multilinear regression.

Chapter 6

FACTORS INFLUENCING THE ASSESSMENT OF ROAD ROUGHNESS

Section 6.1 Introduction

This Chapter addresses Objective 3 of the research, via a comprehensive literature review allied to a numerical modelling study which identify and compare the factors affecting vehicle vertical acceleration responses, measured by a smartphone mounted in a vehicle moving at normal traffic speeds, to a road profile. Moreover, the chapter describes a framework to quantify the effects of different factors on the measured vehicle response, which can be used to develop a road roughness measurement prototype system. For example, a smartphone-based system that is capable of use by the public/stakeholders to acquire measurements of road roughness at the network level.

The methodology developed to investigate the relative effects of the different factors affecting the measurement of vertical vehicle body acceleration consists of:

- A review of the literature to identify the factors.
- Carrying out computer vehicle simulations to quantify the effects of these factors
- Discussing the findings from simulations together with empirical studies to summarise the relative importance of different factors (i.e. high, medium or low)
- Developing algorithms to improve the understanding of the relative importance of the factors and to gain an understanding of the levels of accuracy these types of

algorithms trialled can achieve.

The literature suggests that simply correlating road roughness with the vertical acceleration of a vehicle does not provide accurate measurements of road roughness because the vehicle response to the road profile are affected by the a number of factors related to the smartphone, the vehicle and how it is driven and the road profile itself (T. Gillespie, 1981). The main factors identified in the literature which may affect the accurate measurement of vehicle vertical acceleration are (Chugh *et al.*, 2014; ASTM, 2012; T. Gillespie, 1981).

1. Characteristics of smartphones
2. Vehicle related factors
 - a) Vehicle Speed
 - b) Driving behaviours
 - c) Vehicle type
 - d) Suspension stiffness and damping
 - e) Wheelbase
 - f) Sprung mass
 - g) Tyre pressure
3. Road surface profile

This chapter investigates the above-mentioned factors.

Section 6.2 Characteristics of Smartphones

The literature suggests that the following characteristics of a smartphone may be important

when using it to measure vertical acceleration (Belzowski and Ekstrom, 2015; Roadroid, 2013; ASTM, 2009).

- Operating range, resolution and sampling interval
- Accelerometer related errors
- Difference in smartphones when measuring road roughness
- Position of placement in vehicle

Section 6.2.1 Operating range, resolution and sampling interval

According to the *American Society for Testing and Materials E950 standards* (ASTM, 2009), in terms of measuring road longitudinal profile using a vehicle fitted with an accelerometer, the sensitivity of the accelerometer should be enough to accommodate the acceleration levels expected from vehicle bounce motions, typically there are ± 1 g (in addition to gravity).

Accelerometers in smartphones were designed for gaming so that rapid and extreme motions of the smartphone as a result of shaking by the user can be detected. These smartphone accelerometers are normally of good quality, typically with a measuring range of ± 2 g to a resolution of 0.002 m/s^2 (Del Rosario, Redmond and Lovell, 2015). Examples of such smartphones include Samsung S1, S2, S3 and Note 1, Note 2; iPhone 5, 6, 7 and 8. These smartphones therefore evidently meet the requirement of being able to measure to measure accelerations of ± 1 g caused by road roughness (Chandra *et al.*, 2012). Their associated resolution of 0.002 m/s^2 allows accelerations to $1/500^{\text{th}}$ of this amount to be

recorded. This allows the detection of small changes in measured vertical accelerations.

The maximum speed at which a vehicle fitted with a smartphone can travel for data collection should comply with the requirements of classes of road roughness measurement device specified in ASTM E950. The ASTM E950 Classification is based on the maximum longitudinal interval at which a device records roughness. This is a function of both the vehicle speed and the sampling frequency of the smartphone.

Table 6.1 summarises the corresponding vehicle operating speed ranges to satisfy the different classes defined in ASTM E950 (ASTM, 2009). The sampling frequency of a smartphone is typically between 40 to 1000 Hz.

Table 6.1: Vehicle operation speeds required for different smartphone measuring sampling frequencies

Sampling frequencies (Hz)	Time interval (s)	Speed range to satisfy class 1 device (km/h)	Speed range to satisfy class 2 device (km/h)	Speed range to satisfy class 3 device (km/h)
40	0.0250	< 4	4 to 22	22 to 43
80	0.0125	< 7	7 to 43	43 to 86
100	0.0100	< 9	9 to 54	54 to 108
400	0.0025	< 36	36 to 216	216 to 432
1000	0.0010	< 90	90 to 540	540 to 1080

Research however suggests that the operating range of a smartphone should be limited to between 80–120 Hz (Jones and Forslof, 2014). At frequencies lower than 80Hz, the operating speed of a vehicle may need to remain fairly slow to satisfy the requirement for the minimum sampling interval. This would suggest that in slow-moving traffic such as in urban environments, where speed limits are less than 50 km/h, a smartphone used for

roughness measurement would satisfy the requirements of class 2/3 devices according to the ASTM E950 specifications. This limits a sampling interval of not greater than 300 mm on the travelled distance.

Section 6.2.2 Accelerometer errors

It is important to know that sensors in a smartphone can vary in terms of their sample rate, measurement range, resolution to acceleration and sensitivity to gravity or temperature (Roadroid, 2013). These can lead to bias error of measured vehicle accelerations.

Most accelerometers in smartphones are micro-electro mechanistic systems (MEMS) that transform a physical quantity to electrical signals using an integrated circuit (IC). There are various error sources that affect the performance of the accelerometer, the major errors being bias offset, scale error and non-orthogonality of axes (referred to as deterministic errors) (El-Diasty and Pagiatakis, 2008).

The bias offset is a systematic error associated with the offset of the sensor's measurement from its true value. The scale factor refers to the ratio between output electric signal and physical force. The non-orthogonality error results from imperfections of mounting of accelerometers along the three axes during manufacturing. There are other types of error, such as bias varying with time due to inertial temperature changing and bias variance due to noise (i.e. random errors). However, the effects of these errors are generally considered to be relative less compared to the errors mentioned above (El-Diasty and Pagiatakis, 2008; Naranjo, 2008). For example, experiments suggested that under a steady temperature, bias

variance due to noise on iPhone 4 devices was found to be between 0.3 to 0.5 mg, compared to 20 mg caused by deterministic errors.

Figure 6.1 demonstrates the impact of common errors produced by accelerometers in smartphones on the sensor output.

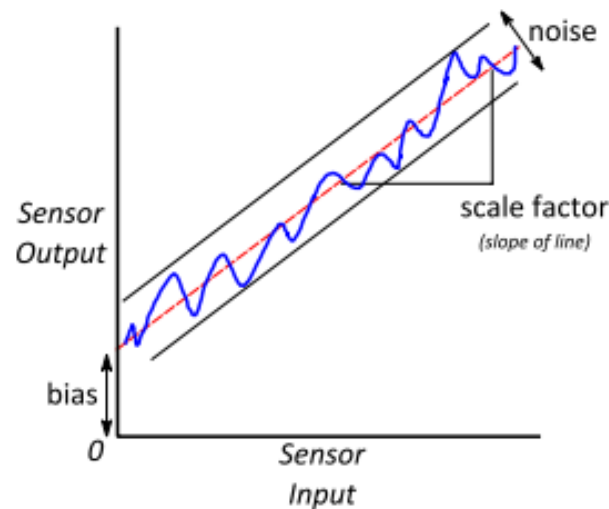


Figure 6.1: Common errors of smartphone accelerometers (NovAtel, 2014)

The accuracy of accelerometers in smartphones is mainly reflected in their deterministic error; most people use ‘bias offset’ as a general term that usually determined from two-position testing (smartphone face-up and face-down test), while precision is associated with the repeatability and reproducibility of using smartphones to capture data (Abdel-Hamid, 2004).

A study was carried out using fourth-generation iPod Touches under steady state to determine the ‘bias offset’ of the accelerometer inside (i.e. an ST MEMS LIS331DL). Five iPod Touch devices were tested, and the error was computed by placing the smartphone face up and face down on the level surface. Results showed that the mean bias offset on

the Z-axis was 0.0165 g (16.5 mg), ranging from 0.0046 g (4.6 mg) to 0.0278 g (27.8 mg) over five devices, with a standard deviation of 0.0086 g (8.6 mg), indicating poor reproducibility. This variance among different devices may be due to variation in physical properties of embedded MEMS, such as misalignment and scale factor error with time.

Kos *et al.*, (2016) carried out a large-scale pilot study of bias measurements of different smartphones. Their experiments consisted of placing the smartphone on a level surface and measuring three-axes accelerations caused by gravity over a period of 100 seconds. 61 types of smartphone were used and in total 500 measurements were taken. The smartphones used in the experiment included the Galaxy S3 and S4; iPhone 4, 5, 5s and 6; Nexus 5 and Xperia Z1 Compact. Their results are shown in **Table 6.2**.

Table 6.2: *Bias measurements of typical smartphone (Kos, Tomažič and Umek, 2016a)*

<i>Smartphone type</i>	<i>Number of phones</i>	<i>Average (g)</i>			<i>Standard Deviation (g)</i>		
		X	Y	Z	X	Y	Z
Xperia Z1 Compact	41	0.0063	0.0279	0.0231	0.0057	0.0019	0.0018
iPhone 4	44	-0.0121	-0.0110	-0.0264	0.0005	0.0010	0.0038
iPhone 5s	42	0.0051	0.0060	0.0034	0.0009	0.0008	0.0097
iPhone 6	14	0.0029	0.0032	0.0004	0.0025	0.0009	0.0008
LG Nexus 5	11	-0.0241	-0.0259	-0.0994	0.0006	0.0017	0.0013

After recalculate the results of experiments, an average of 0.025 g (25 mg) bias offset was found in the vertical Z direction for all smartphones. This suggested that the effect of bias offset on the measurement of vehicle body acceleration cannot be ignored particularly when the vehicle body accelerations measured to vary between 0.07 to 0.7 g on a state roads (Chandra *et al.*, 2012).

Generally, more recent generations of smartphones tend to be more accurate (Buttlar and Islam, 2014). As can be seen from **Table 6.2**, for example, the iPhone 6 has an average bias offset of 0.0004 g (0.4 mg) compared to the iPhone 4 of 0.0264 g (26.4 mg). Besides, the smaller standard deviation for the iPhone 6 also suggested reproducibility of results.

The above findings are also confirmed by shaking table tests carried out by Feng et al., 2015. In their experiment, three popular smartphones (see **Table 6.3**) were subjected to sinusoidal motions with accelerations ranging from 0.05 to 0.2 g. These results showed that more recent smartphone accelerometers (iPhone 5) were more accurate in the measurement than older devices (iPhone 3), the average error rate for iPhone 3 is 21.9 % whereas 5.6% for iPhone 5. Specifications of accelerometers in the test are listed below (Feng *et al.*, 2015):

Table 6.3: *Specifications of accelerometers used in shaking table test* (Feng *et al.*, 2015):

<i>Property</i>	<i>Reference</i>	<i>Smartphone 1</i>	<i>Smartphone 2</i>	<i>Smartphone 3</i>
Sensor maker	PCB Piezotronics	ST Micro-electronics	ST Micro-electronics	Bosch Sensortec
Sensor model	393B04	LI331DL	LIS331DLH	SMB380
Phone maker & Operating Sytem (OS)	NI SCXI-1531	iPhone 3GS, iOS	iPhone 5, iOS	Samsung Galaxy S4, Android
Type	Piezoelectric	MEMS	MEMS	MEMS
Measurement range	5g	2g	2g	2g
Output data rate/ frequency range	0.05 to 700 Hz	100 to 400 Hz	0.5 to 1000 Hz	0.5 to 3000 Hz
Noise density (mg/ \sqrt{Hz})	0.0004	N/A	0.218	0.5

In addition, not only can the above issues related to smartphone type cause errors, but also external factors can result in a random error of the accelerometer. One of the factors with the greatest impact is the temperature which affect the performance of the smartphone's micro-electro mechanistic systems (MEMS). MEMS based accelerometers are normally more temperature sensitive than other non-MEMS sensors due to their miniature and lightweight design, particularly with respect to deterministic errors including bias and scaling factor error (Abdel-Hamid, 2004). The MEMS converts the physical accelerating force to an electric charge. The accuracy of the DC component of the charge is temperature dependent and as well as to the component's lifetime (Aljeroudi, Legowo and Sulaeman, 2006). Increases in smartphone temperature are predominantly caused by the operation of the Central process unit, GPS module and battery recharging. These inertial temperature fluctuations cause a large bias drift in measured accelerations. At room temperature however, the ambient air temperature generally has less effect on the changes of temperature in smartphones (Kos, Tomažič and Umek, 2016b).

A study was carried out by Abdel-Hamid (2004) to investigate the temperature sensitivity of accelerometers, of a similar type to those used in a typical smartphone, when measuring gravity under steady state conditions using a climatic chamber. In these experiments, the temperature was gradually raised from -30 °C to 30 °C in increments of 5°C. The results showed that measured accelerations generally increased with temperature (see **Figure 6.2**). and that the associated bias drift caused by temperature was approximately 0.7 mg/°C.

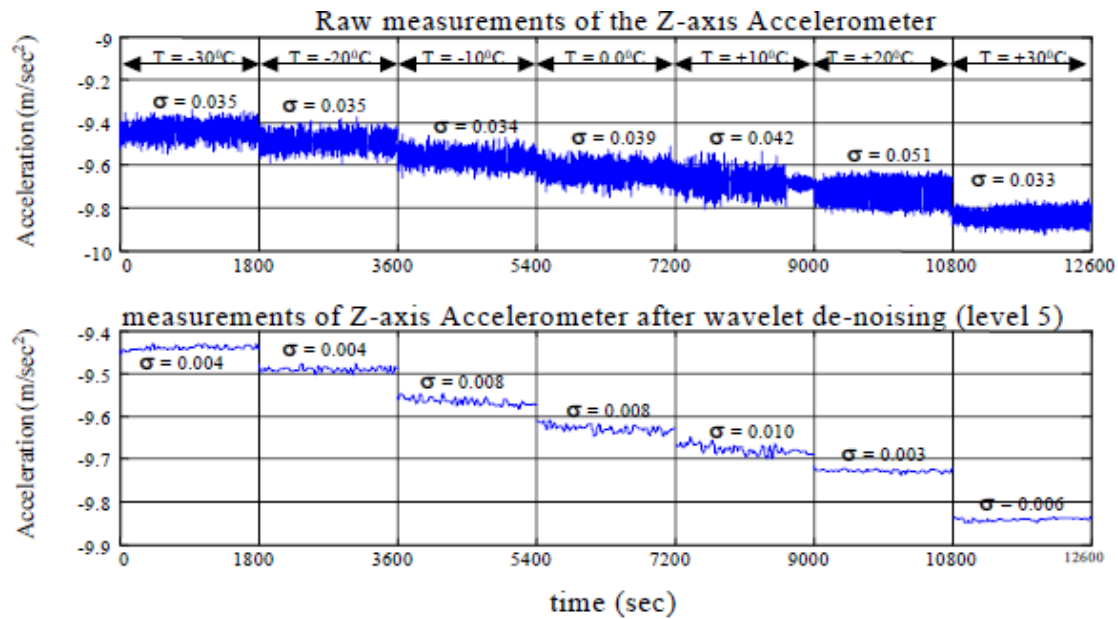


Figure 6.2: Effects of temperature and time on measurements of accelerations using a smartphone (Abdel-Hamid, 2004)

Kionix (2015) determined the temperature sensitivity of a number of different types of accelerometers typically found in smartphones. He found that an MEMS accelerometer can have an offset ranging from 1 to 2.4 $\text{mg}_0/^\circ\text{C}$ caused by temperature changes (Kionix, 2015). For example, the accelerometer devices ADXL330 and ADXL 345 were found to have temperature offsets of 2.3 $\text{mg}/^\circ\text{C}$ and 1.2 $\text{mg}_0/^\circ\text{C}$, respectively (Analog Devices, 2018; Naranjo, 2008). Accordingly, the above would suggest that increased temperature of 10 $^\circ\text{C}$ may result in 7 to 24 mg of temperature offset.

Section 6.2.3 Differences in smartphones when measuring IRI

From the above it may be seen that the performance of accelerometers in smartphones can vary due to bias offset and manufacturing issues and temperature sensitivity or random noise. The issues can lead to variances when collecting vehicle body accelerations caused

by road roughness using mounted smartphones, resulting in discrepancies of the predicted IRI values.

A number of researchers have demonstrated the differences in measured road roughness using different smartphones. Douangphachanh and Oneyama (2013) found that there was a difference between the Samsung Galaxy Note2 and S3 in terms of the average IRI measurement when the smartphones were mounted in the same vehicle. Belzowski and Ekstrom (2015) reported that smartphones themselves were found to be a significant factor affecting the road IRI prediction through correlation of measured vehicle responses.

To gain further insight into the acceleration data obtained by different smartphones, this research investigated the use of two smartphones (iPhone 5 & 7) fitted in the same position within a BMW M4 sports car (see **Figure 6.3**). Data was captured using the two devices over a 1 km length of the B41144, in Star City, Birmingham. Acceleration data were recorded with a Sensorlog smartphone APP (Thomas, 2018) (this can be freely downloaded). The sampling frequency of both devices was set to 100 Hz resulting in 3971 measurements. The Root Mean Square (RMS) of the measured vehicle body acceleration of the two smartphones were used to make the comparisons. The results showed that, generally, the measured values of the iPhone 5 (i.e. 0.13 g) were larger than those measured with the iPhone 7 (i.e. 0.10 g), about 0.025 g on average. A paired sample T-test also showed that two measured samples are statistically different (at the 95 % confidence level) (as shown in **Figure 6.4**). The difference between RMS values ranged from 0.02 to 0.028 g.



Figure 6.3: Car used for testing (left) and mounting position of smartphones (right)

➔ **T-Test**

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	i7	-.922605124	3971	.0728155546	.0011555114
	i5	-.914153572	3971	.0937619189	.0014879096

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	i7 & i5	3971	.051	.001

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	i7 - i5	-.008451552	.1157231072	.0018364121	-.012051951	-.004851152	-4.602	3970	.000

Figure 6.4: Results of paired sample T-test

The reasons for these differences might be attributed to: a) a scale factor error related to the non-orthogonality of the iPhone 5 subject to dynamic conditions. The iPhone 5 was three years old and been dropped on a number of occasions, while the iPhone 7 was a brand new device; and/or b) significant random noise causing errors in the iPhone 5.

Apart from the above issues associated with the accelerometers inside smartphones, the measurement of IRI accurately relies on accurate distance measurements. Most smartphones have an inertial satellite navigation GPS system, which enables the travelled distance and vehicle speeds to be measured. However, the accuracy of the GPS in measuring travelled distance determines the accuracy of measured road IRI.

An error of $\pm 3.3\text{m}$ was found in the accuracy of smartphone GPS positioning in field experiments carried out by Eriksson et al., (2008). GPS accuracy in HTC Desire and Samsung Galaxy S smartphones were reported to be $\pm 6.3\text{ m}$, respectively (Strazdins *et al.*, 2011). Jang, Yang et al. (2016) also noted that the positioning error of mobile phone data collection is typically less than 3m when communication between the GPS detector and satellites is not significantly blocked.

Section 6.2.4 Position of the Smartphone

The position in which smartphones are mounted inside a vehicle has also been found to affect the measured vehicle body acceleration. Research by Kropáč and Můčka (2005) is useful in trying to quantify what the effects might be. They built a 12 degrees of freedom planar stimulation of a travelling passenger car to calculate the RMS vertical acceleration of driver and passenger seats (as shown in **Figure 6.5**).

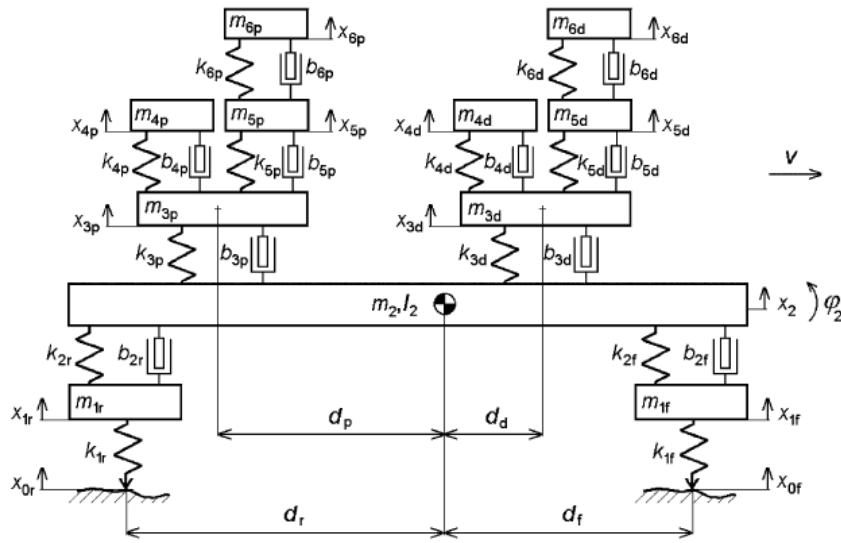


Figure 6.5: 2D stimulation model for passenger cars (Kropáč and Múčka, 2005)

Kropáč and Múčka (2005) showed using their model that the average RMS accelerations of a passenger seat was about 15% larger than those measured by the driver seat. Their finding is very similar to another study which simulated a Land Rover Defender 110 in vehicle dynamic simulating package (Uys, Els and Thoreson, 2007). Uys *et al.*, (2007) found that the difference of RMS acceleration of driver and passenger seats was 13%. Douangphachanh and Oneyama (2014) also found, in an experiment of using smartphone to measure road roughness, that the different locations of smartphone mounted inside the vehicle can affect the accuracy of road roughness.

The acceleration readings recorded by smartphone accelerometers can also be sensitive to the mounting arrangement in the vehicle. Belzowski and Ekstrom (2015) suggest that the smartphone should be mounted on the windscreen with a suction cup, with a bracket that attaches to the dashboard to ensure a rigid mounting. They also postulated that when a

smartphone is mounted on the windscreen without being attached to the dashboard, it was compelled to ‘float’ if it was not firmly mounted between the dashboard and windscreen (Belzowski and Ekstrom, 2015). Accordingly, the readings of the signal would lack fidelity due to the flexibility of the plastic bracket. However, by comparing the acceleration signals measured by the above-mentioned two mounting methods to measured vehicle accelerations caused by roughness, no significant difference was found (as shown in **Figure 6.6**) (Eriksson *et al.*, 2008).

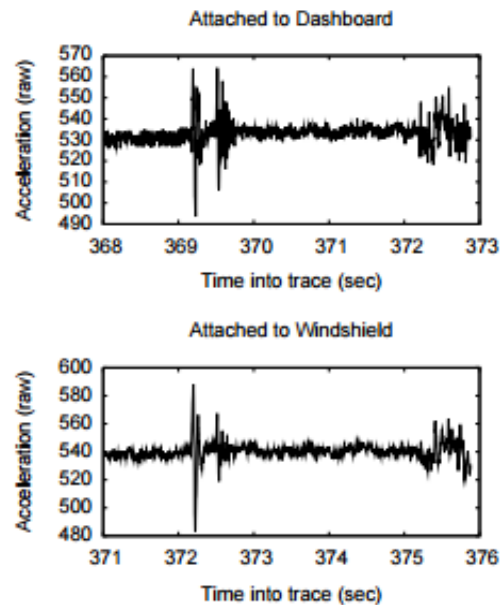


Figure 6.6: *Comparison of signals recoded from two positions of accelerometers mounted in the vehicle* (Eriksson *et al.*, 2008)

The angle of orientation of the smartphone, is also an important consideration. **Figure 6.7** depicts the six positions in which one of the axes of the three acclerometers in a smartphone are parallel to the direction in which gravity acts. The associated acceleration due to gravity for for all six orientations is summarised in **Table 6.4**.

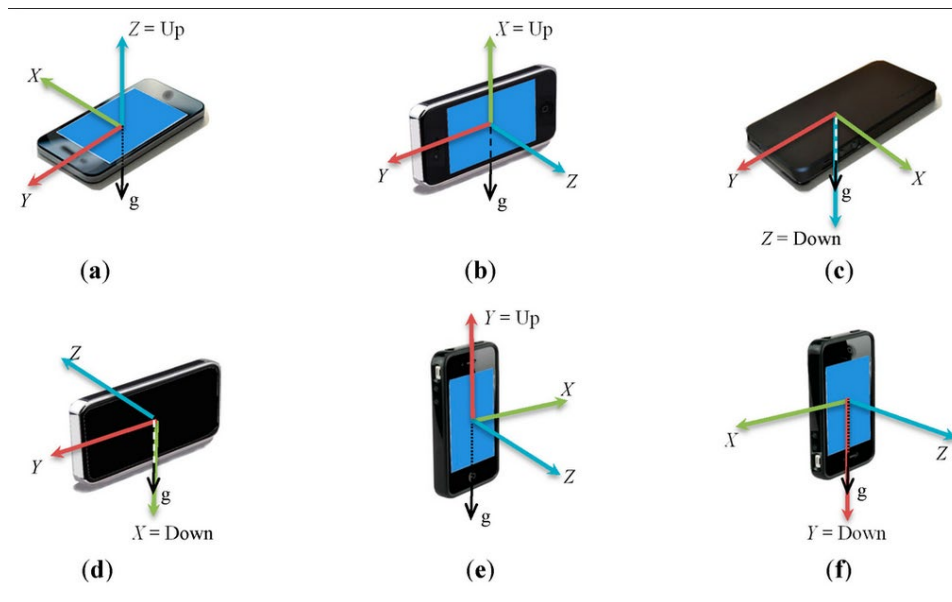


Figure 6.7: Six positions where smartphone devices are normally mounted in the vehicle (Saeedi and El-Sheimy, 2015).

Table 6.4: The effects of different smartphone orientations (Saeedi and El-Sheimy, 2015).

Position	X-axis	Y-axis	Z-axis	Vertical direction
A	0	0	-9.81m/s^2 (g)	Z
B	-9.81m/s^2 (g)	0	0	X
C	0	0	9.81m/s^2 (g)	Z
D	9.81m/s^2	0	0	X
E	0	-9.81m/s^2	0	Y
F	0	9.81m/s^2	0	Y

As may be seen from **Table 6.4** when the smartphone is mounted in a position other than these ideal orientations the direction of the acceleration due to gravity is not parallel resulting in an associated change in the gravity associated acceleration in the x, y and z directions. See **Figure 6.8** and **Table 6.5**.

The most common position of mounting a smartphone in the vehicle using brackets are

positions (b) and (e), where the driver is able to view and operate the smartphone before move off. Given the above, it is important to ensure that the smartphone is as close as possible in one the positions shown in **Figure 6.7** because of the need to measure vehicle vertical body accelerations caused by road roughness. Alternatively, if the angles of inclination are known the Pythagorean theorem may be used to infer the acceleration in the vertical direction from the x, y and z smartphone accerlerations.

Table 6.5: *Effect of angle of smartphone orientation on measured gravity*

<i>Angle θ (°)</i>	<i>a_y (g)</i>	<i>a_z (g)</i>
50	-0.766	-0.643
60	-0.866	-0.500
70	-0.940	-0.342
80	-0.985	-0.174
85	-0.996	-0.08
90	-1.000	0

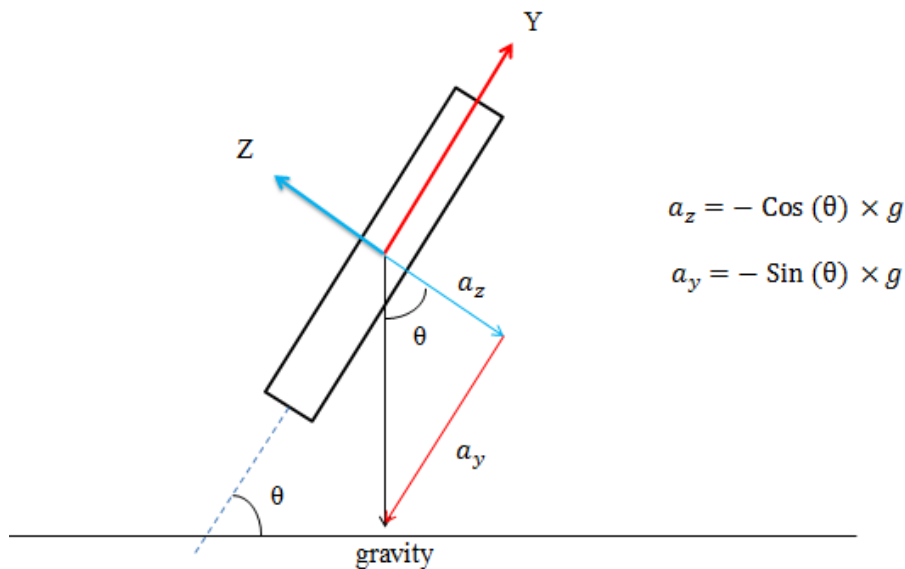


Figure 6.8: *Measured accelerations in y and z axes*

Belzowski and Ekstrom (2015) found that although the securely fastening the smartphone and accounting for its orientation are important factors, the effects of different methods of mounting a smartphone are significantly less.

In Section 6.2, the characteristics of smartphones on measuring vehicle vertical body acceleration were investigated. **Table 6.6** provides a summary of the smartphone related factors discussed above.

Table 6.6: *Smartphone specification needed to measure road roughness*

Sepcifications	Details
Sampling freqnecy	Should able to achieve 80Hz to 120Hz
Measurement Range	At least $\pm 2g$
Revolution	Recommended to be equal to or less than 0.002 m/s^2
Temperature sensitivity	less than $2.4 \text{ mg/}^\circ\text{C}$
Bias error	Less than 0.02 g
Recommended smartphone models	iPhone 5 and later version or equivalent models that were manufactured after 2010, such as Motorola Droid X2 or better smartphone coming later

Section 6.3 Vehicle related factors

Following a review of the literature, **Figure 6.9** summarises the main vehicle related factors that influence body acceleration measurement using a smartphone (ASTM, 2012).

To further understand and quantify how the road roughness measurement is affected by

vehicle related factors, a parametric study was carried out based on a suite of numerical simulations using the CarSim™ vehicle dynamic (Mechanic Simulation, 2017).

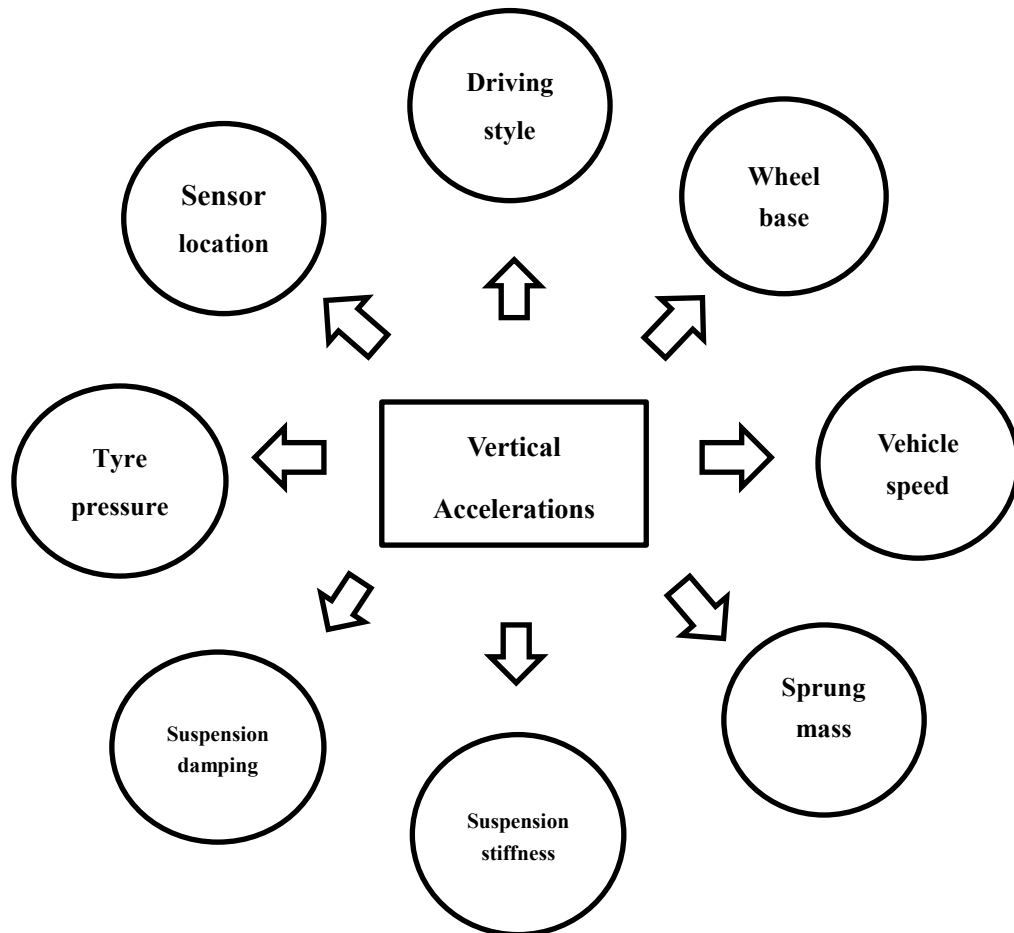


Figure 6.9: Factors that influence the vertical accelerations collected using Smartphone data collector

CarSim™ vehicle dynamic package

The CarSim™, is an industry standard commercial software that has been extensively used as a vehicle dynamics simulator (Benekohal and Treiterer, 1988; So *et al.*, 2014; Mechanic Simulation, 2017). It provides a 3D non-linear motion of rigid vehicle bodies, taking into account the effects of the kinematics and compliance on the suspension and steering. It is

used to simulate vehicle dynamics (e.g. steering, accelerate and brake, etc.) in response to driver control under a given environment (road geometry, friction, wind, etc.).

Simulations

The simulation carried out consisted of determining the vertical vehicle body accelerations for a variety of combinations of vehicle type, speed, sprung mass, suspension and tyre pressure, driving behaviour and road profile. Throughout it was assumed that a vehicle body acceleration was equivalent to the vertical acceleration measured by a smartphone. The road profiles were straight sections of road of 160 m in length, with IRI values of between 1.0 m/km and 7.3 m/km to represent road of conditions ranging from very good to poor. The modelled road profiles were acquired from actual roads provided in the LTTP dataset (FHWA, 2018). For all simulations, a 70kg extra weight was added to the vehicle's sprung mass to take in account the weight of the driver.

The literature suggests that the power spectral density (PSD) of a vehicle's body acceleration, acquired from analysis in frequency domain, correlates well with road roughness (Marcondes *et al.*, 1992; Hesami and McManus, 2009; Sun, 2001). Parseval's Theorem states that the Root-Mean-Square of the vehicle body vertical acceleration (Grms) in the time domain is an equivalent measure to the square root of the integral of the PSD (Van Baren, 2012; Simmons, 1998; Rogers *et al.*, 1997; Doertenbach, 2012). Therefore, for the purpose of the analysis the relative effect of different vehicle related factors on the measured vehicle vertical accelerations, the Grms values of vehicle body accelerations acquired from the CarSimTM simulations were compared with the road

profile IRI values (Jang *et al.*, 2016; Dawkins *et al.*, 2011). The Grms can be calculated as follows:

$$Grms = \sqrt{\frac{1}{N} \sum_{i=1}^N (a_{z,i} - g)^2}$$

Equation 6.1

Where $a_{z,i}$ = the i^{th} vehicle body vertical accelerations (g);

N = the number of acceleration readings recorded over the measured road section;

g = Gravity (ms^{-2});

Section 6.3.1 Vehicle Speed

Research has shown that the measured vertical vehicle acceleration caused by road roughness is related to vehicle speed (Douangphachanh and Oneyama, 2014; Zeng, 2014; Du *et al.*, 2014). Theoretically, considering a simple sine wave modelling of road roughness, if the road roughness is considered as an input to the vehicle wheels, causing vehicle vibrations as a function of travel speed, then, for a given temporal frequency, the amplitude of acceleration inputs increases with the square of the vehicle speed (T. D. Gillespie, 1992a).

To better understand the influence of vehicle speed on the measured vehicle body acceleration, a D-class Sedan was modelled in CarSimTM. A D-class Sedan is equivalent to an Audi A1. The speed of the modelled vehicle ranged from 30 km/h to 100 km/h. Four

road sections were considered with corresponding IRI values of 1.0, 3.1, 5 and 7.3 respectively. The resulting Grms values of vehicle vertical body acceleration are shown in *Figure 6.10*.

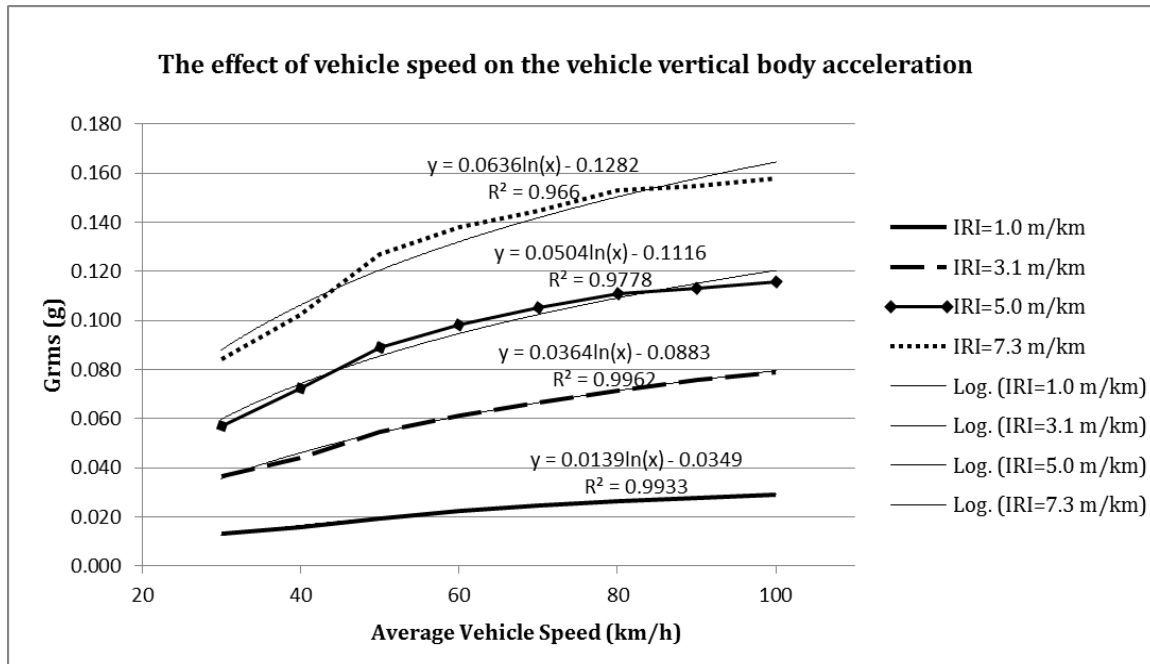


Figure 6.10: The effect of vehicle speed on vehicle vertical body accelerations

As can be seen from the results, increasing vehicle speed resulted in a logarithmic increase in the measured vehicle body acceleration (Grms). The increase of Grms was more prominent at higher road roughness value. These results corroborates empirical studies showing that the PSD value of both vehicle body and wheel acceleration increase with vehicle speed (Gao and Chen, 2011; González *et al.*, 2008; Levulytė, Žuraulis and Sokolovskij, 2014; Du *et al.*, 2014). It was also found that the vehicle body acceleration increased on average by 93% over all the road sections considered when speed increased from 30 km/h to 100 km/h.

The results also show that the increase in Grms tends to reduce at vehicle speeds above 80 km/h. Since a road, mostly consisted of many “short-wave” defects, e.g. cracking and potholes. By increasing vehicle speed, vibrations caused by defects are generally eliminated by car suspension particularly at speeds greater than 80 km/h (as shown in **Figure 6.11**).

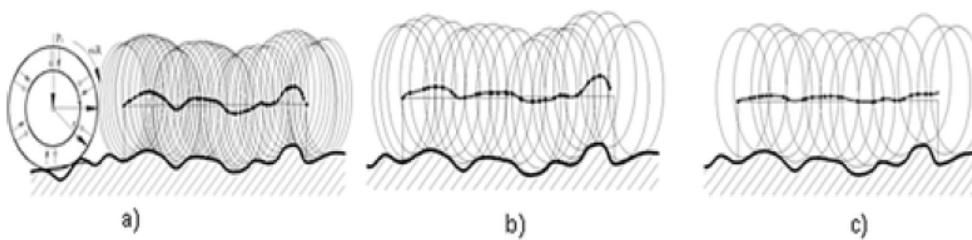


Figure 6.11: Vehicle wheel displacement model for different vehicle driving speeds: a) low speed < 20km/h; b) normal speed (20 to 80km/h); c) high speed, e.g. > 80km/h
(Levulytė *et al.*, 2014)

Section 6.3.2 Vehicle type

The influence of vehicle type on the measured vehicle body acceleration was investigated by undertaking simulations using 15 vehicle types (see. **Table 6.7**). All of the selected vehicle types were modelled travelling at a constant speed of 50 km/h on four 160 m road sections with IRI values of 1.0, 3.1, 5 and 7.3 m/km respectively, representing very good, good, fair and poor conditions. Comparisons were made of the simulated Grms of vehicle body acceleration and results were shown in **Figure 6.12**.

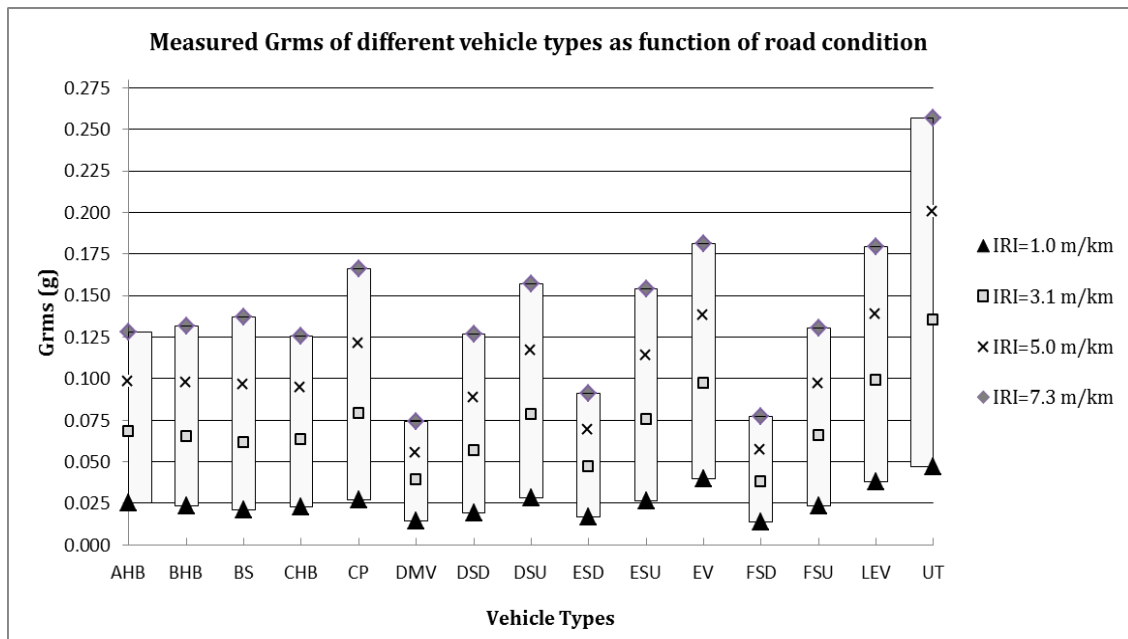


Figure 6.12: Simulated Grms of different vehicle types as a function of road condition

As can be seen from **Figure 6.12**, for all vehicle types considered, the simulated Grms increases with IRI (i.e. with worsening road condition) and the differences in simulated Grms between vehicle types also increases as the road condition worsens.

Unsurprisingly the LEV, EV and UT showed the highest Grms response for all four road conditions. These responses were between 40% and 190% greater than for passenger vehicles, depending on road condition, reflecting the relatively stiff suspension and heavy un-sprung masses of vans and utility trucks commensurate with their requirement to carry cargo. As the differences in simulated Grms for vans and trucks are so much greater than for all other vehicle types, it would be better to exclude these vehicle types from a system which considers data from a fleet of many different classes of vehicles, particularly on roads of relatively poor condition ($IRI > 5$ m/km). Alternatively, a fleet consisting of only these vehicle types could be used. The lowest Grms response, for all four roads, occurred

for DMVs and FSDs. This is due to them being relatively heavy vehicles and for an FSD, which is considered to be a luxury car, its suspension is relatively soft.

Some vehicle types are alike in terms of simulated Grms (within $\pm 10\%$) and will therefore provide similar IRI values for the same road sections. These include Hatchbacks (A, B and CHB) and D-class sedans; D and E classes of Sport Utility Vehicles (DSU and ESU) and; the two types of European vans (EV and LEV). This is because these groups are of a comparable shape, chassis configuration and weight and therefore have similar wheel-hop natural frequencies (González *et al.*, 2008). The acceleration data provided by a fleet consisting of such similar vehicle types fitted with comparable smart phones and travelling at similar speeds could therefore be used to provide reliable estimates of roughness data, to IQL-3/4 without the need to adjust for the vehicle type (see **Table 2.4**).

Vehicle type	Example	Abbr.	Sprung mass (kg)	Length *Width *Height (mm)	Wheelbase (mm)	Front suspension		Rear suspension	Track width (wheel centre) (mm)	Tyre size
						Mass (kg)	Spring rate (N/mm)	Mass (kg)		
Utility Truck	ATV mini truck	UT	600	2750*1260*2750	1923	80	35	88	1260	Pacekka 5.2
Large European Van	Skoda van	LEV	2100	4150*2100*1960	3100	150	98	150	1550	265/75 R16
D-class Minivan	Volkswagen touran	DMV	1800	4150*1905*1669	3000	100	31	100	1640	235/65 R17
European Van	Renault Kangoo	EV	1100	3600*1700*1800	2580	100	31	100	1500	185/65 R15
D-class SUV	Audi Q2	DSU	1430	3725*1845*1705	2660	80	130	100	1565	235/55/R18
E-class SUV	Audi Q3/BMW X1	ESU	1590	4220*1875*1800	2950	120	146	150	1570	265/75/R16
Full-size SUV	Audi Q5/BMW X5	FSU	2257	4475*2029*1955	3140	125	189	150	1725	275/65/R18
Compact Pickup	Nissan Navara	CP	1306	4175*1665*1840	2780	90	130	125	1650	215/70 R15
A-class Hatchback	Smart fortwo	AHB	750	3000*1780*1160	2350	41.5	18	41.5	1415	175/65 R14
B-class Hatchback	Smart forfour	BHB	1100	3300*1695*1535	2600	60	28	60	1480	185/65 R18
B-class Sport car	Mini	BS	1020	3225*1750*1200	2330	60	130.5	60	1480	205/45 R17
C-class Hatchback	Mini	CHB	1270	3850*1916*1610	2910	71	27	71	1675	215/55 R17
D-class sedan	Audi A1	DSD	1370	3975*1795*1471	2780	80	153	80	1550	215/55 R17
E-class sedan	BMW 3 series	ESD	1650	4325*1880*1480	3050	90	24	90	1600	225/60 R18
F-class sedan	BMW5 series	FSD	1820	4500*1870*1475	3160	100	83	100	1605	225/60 R18

Table 6.7: Parameters of different vehicle configurations in CarSimTM

Section 6.3.3 Suspension stiffness and damping ratio

Simulations were carried out to understand the effect of suspension stiffness and damping ratio through the Grms response of a DSD and a LEV. For the comparisons, four 160m long road sections were used with IRI values of 1.0, 3.1, 5.0 and 7.3 m/km respectively, representing very good, good, fair and poor road conditions. The vehicles were modelled as travelling at a speed of 50 km/h. The results are summarised in **Table 6.8**.

For the analysis of suspension stiffness, the vehicle's front and rear suspension spring rates were varied by $\pm 20\%$. The results show that decrease in suspension stiffness (i.e. improving the ride comfort) reduced the simulated Grms values for the DSD by approximately 0-7% and 7-15% for the LEV respectively. On the other hand, by increasing the suspension stiffness, the simulated Grms increased for both vehicle types considered, i.e. 1-7% for the DSD and 8-15% for the LEV respectively. These results correspond with the findings of Gao and Chen, (2011). Such differences were found to be greatest for roads with higher IRI values (relatively worse road conditions). The effect of suspension stiffness on the LEV is generally higher than the DSD because of its relatively stiffer suspension (i.e. with solid rear axle and dependent suspensions) resulting in a higher Grms response.

For the investigation of the suspension damping ratio, the damping ratio of both the front and rear suspension were varied by $\pm 20\%$ and results are shown in **Table 6.8**. As can be seen from the results, by decreasing the suspension damper ratio, the simulated Grms

values were reduced for both vehicle types simulated and these effects increased with worsening road condition. A reduction of 20% in the damping ratio resulted in a reduction in the simulated Grms by 2-7 % for a DSD and 3-8% for a LEV respectively. However, increasing of damping ratio by 20% resulted in an increase of the simulated Grms by 3-7 % for a DSD and 0-8% for a LEV. These effects increased with worsening road condition.

Table 6.8: The effects of suspension stiffness, damping and tyre pressure

Vehicle type	IRI	stiffness ratio	Grms	Differences	Damping ratio	Grms	Differences	Tyre spring rate	Grms	Differences
DSD	1	100%	0.021	-	100%	0.021	-	100%	0.021	-
DSD	1	80%	0.021	-0.49%	80%	0.019	-5.83%	90%	0.020	-3.40%
DSD	1	120%	0.021	1.05%	120%	0.022	6.31%	70%	0.018	-11.17%
DSD	3.1	100%	0.057	-	100%	0.057	-	100%	0.057	-
DSD	3.1	80%	0.082	-5.43%	80%	0.055	-2.81%	90%	0.056	-1.93%
DSD	3.1	120%	0.060	5.96%	120%	0.059	3.68%	70%	0.055	-3.86%
DSD	5	100%	0.087	-	100%	0.087	-	100%	0.087	-
DSD	5	80%	0.082	-5.43%	80%	0.083	-4.39%	90%	0.085	-2.31%
DSD	5	120%	0.092	6.71%	120%	0.091	5.55%	70%	0.082	-4.97%
DSD	7.3	100%	0.115	-	100%	0.115	-	100%	0.115	-
DSD	7.3	80%	0.108	-6.26%	80%	0.112	-6.43%	90%	0.112	-3.13%
DSD	7.3	120%	0.120	4.00%	120%	0.117	4.83%	70%	0.109	-5.39%
LEV	1	100%	0.034	-	100%	0.034	-	100%	0.034	-
LEV	1	80%	0.031	-7.46%	80%	0.031	-7.46%	90%	0.033	-1.49%
LEV	1	120%	0.036	8.06%	120%	0.036	8.06%	70%	0.034	2.09%
LEV	3.1	100%	0.099	-	100%	0.099	-	100%	0.099	-
LEV	3.1	80%	0.087	-12.41%	80%	0.095	-3.94%	90%	0.096	-3.43%
LEV	3.1	120%	0.113	14.13%	120%	0.099	0.30%	70%	0.080	-19.17%
LEV	5	100%	0.146	-	100%	0.146	-	100%	0.087	-
LEV	5	80%	0.130	-11.15%	80%	0.137	-6.33%	90%	0.085	-2.31%
LEV	5	120%	0.164	12.34%	120%	0.156	6.53%	70%	0.082	-4.97%
LEV	7.3	100%	0.149	-	100%	0.149	-	100%	0.115	-
LEV	7.3	80%	0.137	-8.59%	80%	0.144	-3.55%	90%	0.112	-3.13%
LEV	7.3	120%	0.162	8.85%	120%	0.156	4.22%	70%	0.109	-5.39%

Section 6.3.4 Tyre pressure

An examination was carried out of the effects of tyre pressure by comparing the responses of a DSD and a LEV. For the comparisons the vehicles were simulated to travel at a constant speed of 50km/h on four 160 m road sections representative of good, fair, poor and very poor conditions. The effect of under-inflation of tyres on the simulated Grms of vehicle body acceleration was investigated by changing the tyre spring rate of the modelled vehicle to 70% and 90% of its default value. The results are shown in **Table 6.8**. In the analysis of tyre pressure, it was found that by reducing the tyre spring rate, the simulated Grms reduced by 2-12% for a DSD and 1-20% for a LEV respectively depending on the road profile and the amount by which the spring rate was changed. The highest reductions in the simulated Grms were found to occur for worse road conditions and at lower tyre spring rates.

Section 6.3.5 Sprung mass

The effect of additional weight to a vehicle (e.g. cargo and passengers) on the simulated Grms of vehicle body acceleration was simulated by changing the vehicle sprung mass. The average weight of one adult in Europe is approximately 70.8 kg (Walpole *et al.*, 2012), and therefore for the purposes of the simulation a 70kg of additional weight was added to the vehicle sprung mass to represent an additional passenger. In the comparisons, five vehicle types were simulated at a travelling speed of 50 km/h on a road in good condition (i.e. IRI=3.1 m/km). The results are shown in **Figure 6.13**.

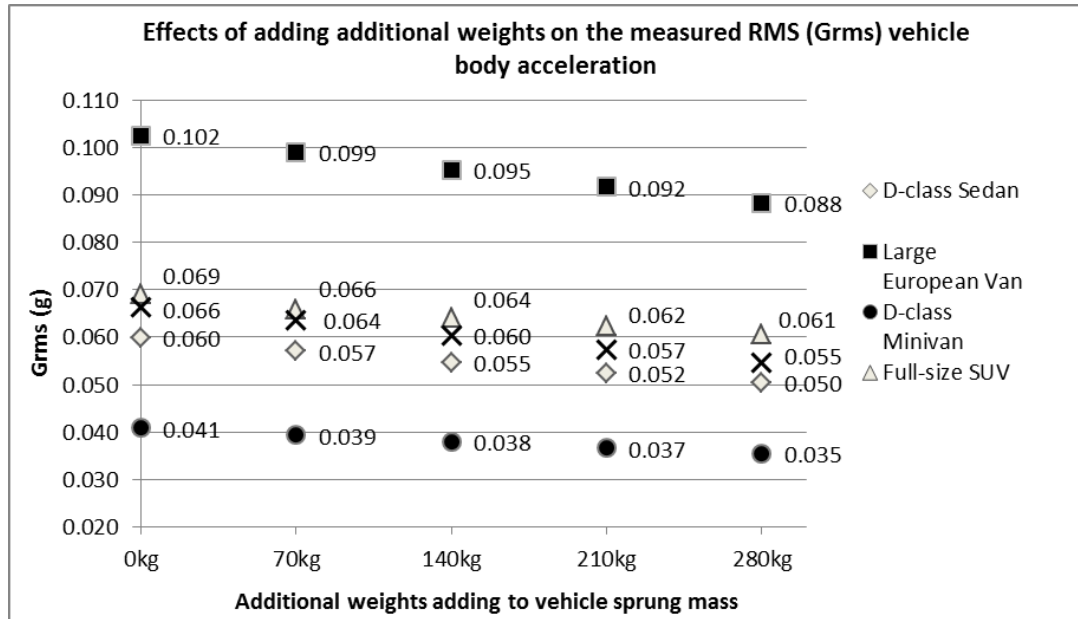


Figure 6.13: The effects of adding additional weight on vehicle sprung mass (Speed=50km/h, IRI=3.1 m/km)

By inspecting **Figure 6.13** it can be seen that adding additional weight to the vehicle sprung mass leads to a reduction in the simulated Grms of the vehicle body acceleration. Every additional passenger (70 kg) resulted in approximately a 5% decrease in the simulated Grms, irrespectively of vehicle type or road condition considered.

Section 6.3.6 Driving style

The effects of different driving style on measured vehicle body acceleration was undertaken by simulating a D-class Sedan (DSD) on a road with and IRI=5.0 m/km (i.e. older road in fair condition) travelling at a speed of 50km/h. Comparisons were made between different driving behaviours given in **Table 6.9**.

Table 6.9: Effects of driving regimes on simulated Grms

Acceleration/Braking	Speed ranges (km/h)	Travelled distance (m)	Time (s)	Grms (g)	Difference compared to driving at constant speed
Constant Acceleration (20% throttle control)	0 to 78	160	11.7	0.083	-2.71%
Constant driving speed	49	160	11.7	0.085	–
Half-throttle accel.	0 to 100	160	10.0	0.087	-6.88%
Constant driving speed	58	160	10.0	0.092	–
Deceleration at constant brake pressure	78 to 0	152	15.9	0.070	4.95%
Constant driving speed	35	152	15.9	0.067	–
Constant speed 100km/h and braking after 2s	100 to 0	137	7.7	0.091	-5.96%
Constant driving speed	64	137	7.7	0.096	–
Full throttle accel, then braking	0 to 82, then to 62	160	10.7	0.085	-5.55%
Constant driving speed	54	160	10.7	0.090	–

As can be seen from **Table 6.9**, compared a constant speed of driving speed, accelerating or decelerating can affect the simulated Grms of vehicle body acceleration by only between 3% to 7%.

Section 6.3.7 Wheelbase

The influence of changing the wheelbase (i.e. distance between the centre of front and rear axles) of a vehicle on the simulated Grms was examined by varying the wheelbase of the four vehicle types shown in **Table 6.10** by ± 300 mm. The simulations were carried out

with vehicles travelling on a road section in good condition (i.e. IRI=3.1 m/km) at a speed of 50km/h. The results are presented in **Table 6.10** from which it may be seen that the differences in the simulated Grms were less than 11% for all vehicle types considered.

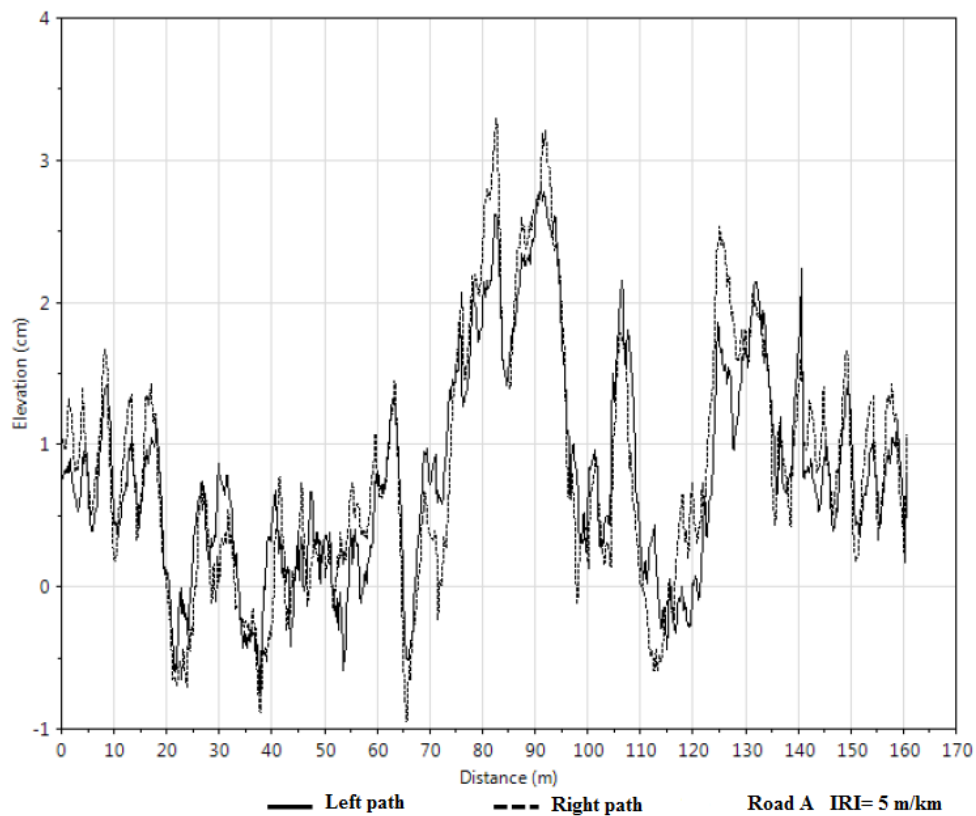
Table 6.10: The effects of wheelbase in the roughness measurement

Vehicle types	Wheelbase changes (mm)	Grms	Change (%) compared with original Grms
D-class sedan	+300	0.059	3.68%
D-class sedan	-300	0.054	-4.90%
Large European Van	+300	0.088	-10.66%
Large European Van	-300	0.101	2.48%
Full size SUV	+300	0.063	-4.63%
Full size SUV	-300	0.068	3.06%
D-class Minivan	+300	0.039	-0.48%
D-class Minivan	-300	0.038	-2.93%

Section 6.4 Effects of longitudinal profile

Road longitudinal profiles of different shapes (i.e. reflected in road elevation PSD) can however have the same IRI value. Although the elevation PSD of every road is unique the measured IRI of road sections can be similar, as IRI is mainly used to indicate a measure of average ride quality over a distance of 160 m and therefore it does not contain any information about the shapes of road sections. An example is shown in **Figure 6.14** of two very different real road profiles (road A and B) that have the same IRI values (i.e. IRI=5 m/km measured over a length of 160m) but with different PSD curves (see **Figure 6.15**). By inspecting the frequency domain plot of road profiles in **Figure 6.15** it may be seen

that Road A has lower amplitudes for wavelengths smaller than 0.5 m/cycle (short wavelength, possibly suggestive of potholes). Whilst, Road B has higher amplitudes in the vicinity of 20 m/cycle wavelengths, suggesting higher amplitudes associated with road subgrade deformation or rutting.



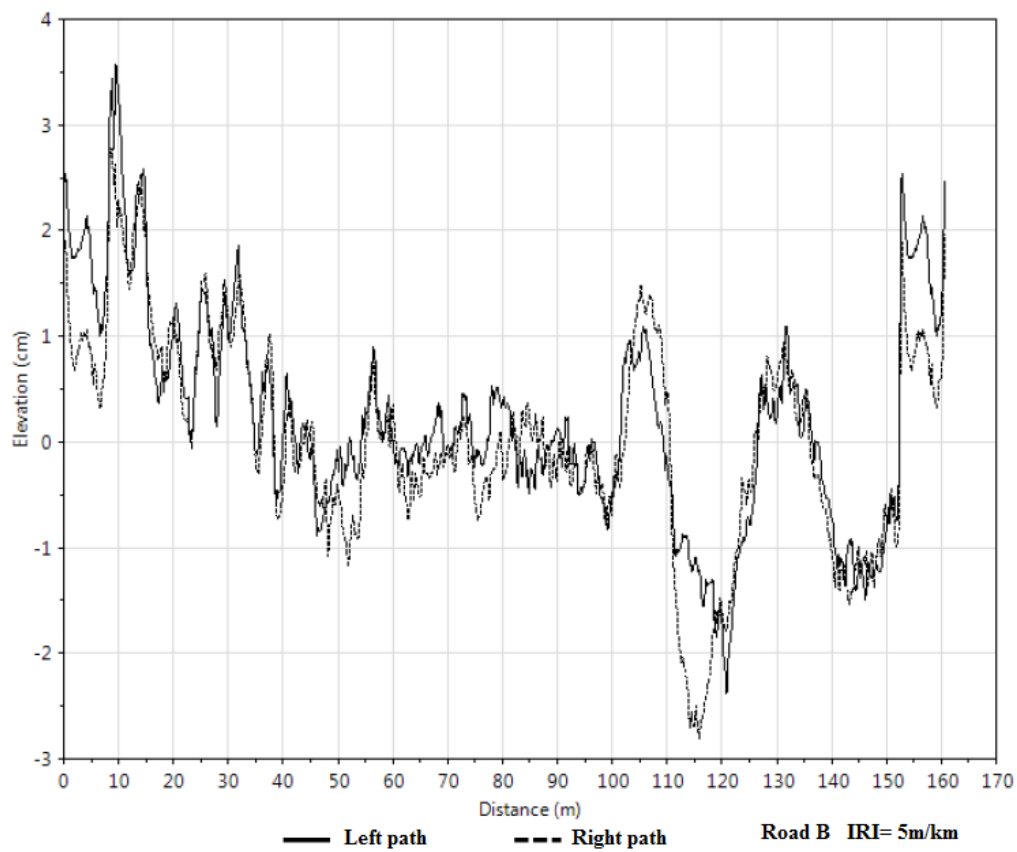
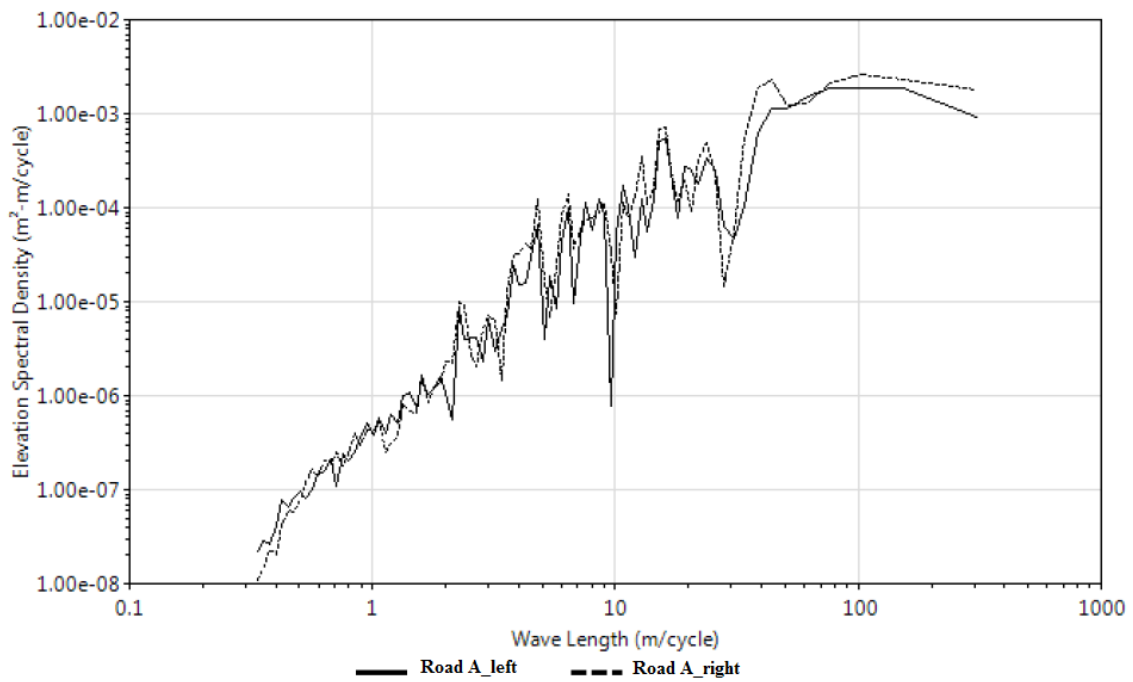


Figure 6.14: Road profiles of two sections (section A and B respectively)



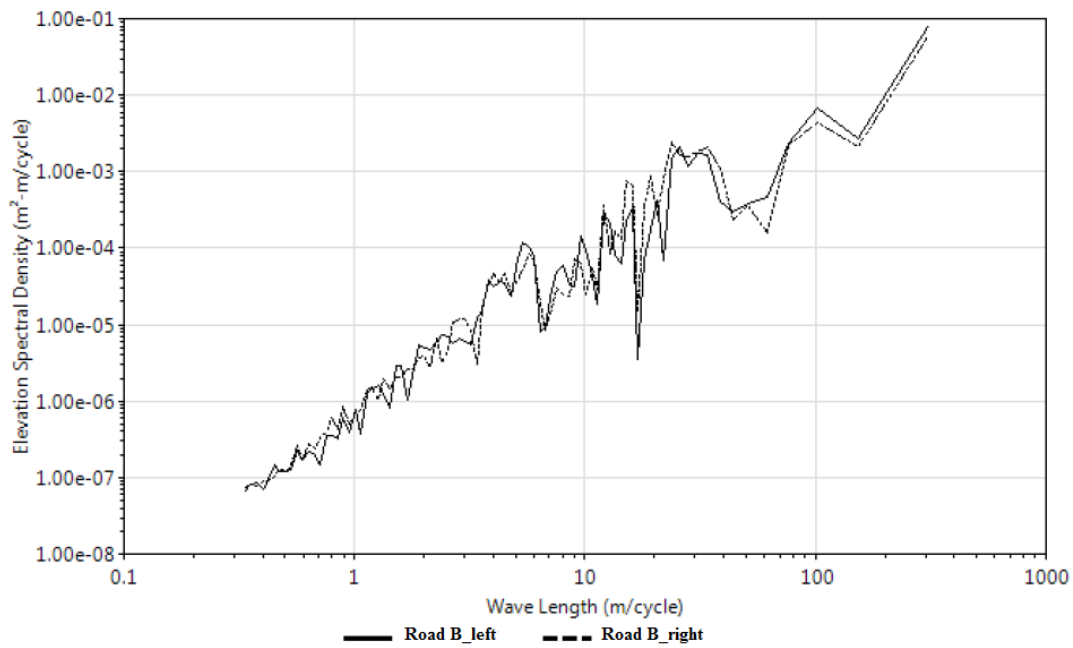


Figure 6.15: The elevation PSD curves for above mentioned two road profiles respectively

Given an overall roughness, in the PSD curves different magnitudes are associated with different wavelengths, which cause the vehicle to oscillate (T. D. Gillespie, 1992a). This implies that the simulated Grms of vehicle body acceleration on two road profiles with same IRI values can also differ.

To further investigate this issue, simulations were undertaken using three vehicle types (i.e. DSD, SUV and EV respectively) travelling on four different road sections (160m in length) at speeds of 30 and 50 km/h. Two of the road sections (A and B) had an IRI of 5 m/km, the other two (C and D) had an IRI of 1.5 m/km. The results are shown in **Table 6.11**. As can be seen from the results, all of the simulated Grms values were different for roads with the same IRI values. Such differences are greatest for a FSU travelling at a speed of 30km/h

on a road section with a low IRI (showing a 25% difference). This implies that it may be necessary to introduce a filter which enables the effects of short wavelength features (i.e. macrotexture, cracks and potholes) on the road profile to be eliminated when assessing, using the advocated approach, roads with low IRI values.

Table 6.11: Grms values for four different road profiles

Vehicle type	Speed (km/h)	IRI = 5 m/km			IRI = 1.5 m/km		
		Road A Grms (g)	Road B Grms (g)	Change (%)	Road C Grms (g)	Road D Grms (g)	Change (%)
D-Sedan	30	0.051	0.053	2.12%	0.015	0.018	19.38%
	50	0.070	0.065	-8.08%	0.021	0.024	14.58%
Full-size SUV	30	0.057	0.061	7.59%	0.017	0.021	24.92%
	50	0.079	0.076	-3.94%	0.023	0.028	23.31%
European Van	30	0.111	0.110	-1.14%	0.047	0.043	-8.52%
	50	0.126	0.109	-12.97%	0.061	0.065	5.62%

Section 6.5 Summary of all factors

Table 6.12 summarises the influence of the factors considered above and classifies the influence into three classes according to their degree of influence (i.e. High >30% Moderate 10-30%, Low <10%).

Table 6.12: The summarised effects of different influencing factors on the measured vehicle body acceleration

Factor	Severity*	Influence
Smartphone Types	H	The literature suggests that different smartphones will give different results when measuring IRI.
Vehicle speeds	H	Simulated Grms increases logarithmically with speed. Increase of 93% in Grms with a speed increase of 266 % (i.e. from 30 to 80 km/h)
Vehicle types	H	Differences in Grms between vehicle types of up to 190%
Sprung Mass	M	Every 70 kg (i.e. approximately one additional passenger) added to the sprung mass results in approximately a 5% decrease in Grms
Longitudinal profile	M	Up to 25% change in Grms due to change in longitudinal profile
Suspension stiffness	M	5 - 15 % change in simulated Grms as a result of $\pm 20\%$ of changes in spring rate
Tyre pressure	M	Up to 20% change in simulated Grms due $\pm 30\%$ change in tyre pressure
Accelerating/Braking	L	3-7% change in simulated Grms due to changing driving behaviour
Suspension damping	L	Less than 8% change in simulated Grms as a result of $\pm 20\%$ of changes in damping
Position inside of vehicle	L	Position should be fixed
Wheelbase	L	Up to 11% of variance

*(High, M: Moderate, L: Low)

Section 6.6 Model Development

6,000 CarSim™ simulations were carried out to obtain a data set which relates IRI to simulated Grms as a function of those factors described above which were found to have a high or moderate impact on Grms (see. **Table 6.12**). Multivariable linear regression and machine learning algorithms were used to analyse the data and develop models to predict IRI from simulated Grms values. The input variables considered in the simulations are listed in **Table 6.13**.

Table 6.13: The values selected for the model development

Variable	Abbreviation	Range of variables considered
Vehicle speed	Speed	30, 50, 80 km/h
Vehicle Type	-	LEV, DSD
Sprung mass (number of people)	Npeop	1, 2, 3, 4 (70kg per person)
Suspension stiffness	Stif	0.8, 1, 1.2
Suspension damping	DampF	0.8, 1, 1.2
Tyre pressure	TyreS	0.7, 0.9, 1
Road profile (IRI values calculated over 160 m)*		1, 2.12, 3.11, 4.24, 4.91, 6.1, 7.42, 9, 12.2 m/km

* Obtained from data available in the LTTTP database (FHWA, 2018)

Table 6.14: Regression Analysis for Large European van and D-class sedan

D-Class Sedan				Commonality		Pearson's Correlation					
Model	Estimated Coefficients β	Coefficient of determination (R^2)	Std.Error	Coefficient (Unique)		IRI	Grms	Speed	Npeo	Stif	Damp F
(Constant)	7.66			-	IRI	1.00	.81	-.05	.00	.00	.01
Grms	55.25			0.82	Grms	.81	1.00	.37	-.06	.15	.07
Speed	-.07			0.14	Speed	-.05	.37	1.00	.01	.01	.01
Npeop	.19	0.83	1.39	0.00	Npeo	.00	-.06	.01	1.00	.00	.00
Stif	-2.93			0.02	Stif	.00	.15	.01	.00	1.00	.01
DampF	-1.09			0.00	DampF	.01	.07	.01	.00	.01	1.00
TyreS	-1.44			0.00	TyreS	.00	.06	.01	-.00	.01	.00

Large European Van				Commonality		Pearson's Correlation					
Model	Estimated Coefficients β	R^2	Std.Error	Coefficient (Unique)		IRI	Grms	Speed	Npeo	Stif	Damp F
(Constant)	5.75			-	IRI	1.00	.83	-.03	.00	.02	-.01
Grms	53.43	0.79	1.54	0.79	Grms	.83	1.00	.33	-.06	.06	.04
Speed	-.06			0.10	Speed	-.03	.33	1.00	.00	.00	-.00
Npeop	.18			0.00	Npeo	.00	-.06	.00	1.00	-.00	.00
Stif	-.87			0.00	Stif	.02	.06	.00	-.00	1.00	.01
DampF	-.89			0.00	DampF	-.01	.04	-.00	.00	.01	1.00
TyreS	-.27			0.00	TyreS	.00	.01	.00	-.00	.00	.00

The results of the multivariable linear regression for the Large European Van and D-class Sedan are shown in **Table 6.14**.

The data in **Table 6.14**, was used to develop three linear models for, a Large European van (Equation 6.2), a D-Class Sedan (Equation 6.3) and one which does not include vehicle type (Equation 6.4).

$$IRI = 55.25 * Grms - 0.07 * Speed + 0.19 * Npeop - 2.93 * Stif - 1.09 * DampF - 1.44 \\ * TyreS + 7.66$$

$$R^2 = 0.83 \quad SE = 1.39 \text{ m/km} \quad \text{Equation 6.2}$$

$$IRI = 53.43 * Grms - 0.06 * Speed + 0.18 * Npeop - 0.87 * Stif - 0.89 * DampF - 0.27 \\ * TyreS + 5.7$$

$$R^2 = 0.79 \quad SE = 1.55 \text{ m/km} \quad \text{Equation 6.3}$$

$$IRI = 50.32 * Grms - 0.06 * Speed + 0.17 * Npeop - 1.86 * Stif - 0.90 * DampF - 0.78 \\ * TyreS + 6.68$$

$$R^2 = 0.74 \quad SE = 1.69 \text{ m/km} \quad \text{Equation 6.4}$$

Where the variables given in the equations are defined in **Table 6.13**.

As can be seen from the results, the coefficient of determination (R^2) of these models are 0.83 for Large European Van (LEV) and 0.79 for D-class sedan (DSD) and 0.74 for model which does not include vehicle type. The standard error of three models are 1.39 m/km,

1.54 m/km and 1.69 m/km respectively. Commonality analysis was carried out to determine how much variance in IRI each of the six variables uniquely contribute. The results, given in **Table 6.14**, show that Grms contributes 82 % of the variance in IRI in the case of the LEV and 79% for the DSD. The Pearson's correlation also shows a strong correlation between Grms and IRI (0.81 and 0.83 respectively). Thus, as might be expected, the simulated Grms is the dominant factor influencing the calculation of IRI.

Speed was found to have the second largest influence on the measurement of IRI, accounting for 14% of the variance in IRI in the case of the DSD and 10 % for the LEV. By ignoring vehicle type, the standard error of regression model has increased to 1.694 (i.e. Equation 6.4), suggesting that vehicle type is an important factor.

Furthermore, it would seem that the other influencing variables do not seem to affect the prediction of IRI significantly. Accordingly, simplified regression models relating IRI to Grms, speed and vehicle type were developed as follows:

For a LEV,

$$IRI = 53.19 * Grms - 0.64 * Speed + 2.99$$

$$R^2 = 0.80 \quad SE = 1.57 \text{ m/km} \quad \text{Equation 6.5}$$

For a DSD,

$$IRI = 52.90 * Grms - 0.055 * Speed + 4.23$$

$$R^2 = 0.78 \quad SE = 1.51 \text{ m/km} \quad \text{Equation 6.6}$$

Ignoring vehicle type,

$$IRI = 49.30 * Grms - 0.055 * Speed + 3.73$$

$$R^2 = 0.73 \quad SE = 1.74 \text{ m/km} \quad \text{Equation 6.7}$$

The values of the R^2 and Standard Error (SE) of the above sets of equations which consider all of the influencing variables, and that which considers speed alone would suggest that multivariate linear regression could be used to determine IRI to a similar degree of accuracy to that which might be expected from a visual inspection (see **Table 2.4**) i.e. IQL-3/4, even without taking it account the class of vehicle. Taking into account all of the influencing variables does not appear to improve the results to an extent which would allow roughness to be measured to a lower IQL.

Machine learning algorithms

A neural network and a decision regression tree machine learning algorithm were developed. These two types of machine learning algorithms were chosen as the literature suggests that they could be suitable for the analysis of relatively small but complex data sets which have a number of independent variables, such as the dataset developed in this study (Melhem and Cheng, 2003; Nitsche *et al.*, 2012; Chandra *et al.*, 2012). Furthermore, their distinctive structure (neuron and tree structure-based respectively) enable them to ‘learn’ the data fast and generate robust predictive algorithms.

To develop the machine learning algorithms, datasets of LEV, DSD and a dataset

combining both vehicle types were partitioned into training (95%) and testing (5%) subsets. The initial analysis of the data was carried out by including all variables defined in **Table 6.13**. Thereafter, a feature elimination process was used to exclude each variable in turn in order to better understand the impacts of the six variables considered on the prediction of IRI. The performances of the algorithms were measured by R-squared and root-mean-squared error (RMSE) between predicted and actual IRI. The results are shown in **Table 6.15**.

Table 6.15: Results of machine learning

Large European Van (LEV)						
Eliminated variable	Neural Networks		Relative Importance	Decision regression tree		Relative Importance
	R-squared	RMSE (m/km)		R-squared	RMSE (m/km)	
None	0.97	0.6	-	0.90	1.0	-
Grms	-0.00	3.4	1	-0.29	3.4	1
Speed	0.70	1.9	2	0.37	2.4	2
N of People	0.96	0.6	5	0.88	1.0	6
Stiffness	0.93	0.9	3	0.76	1.5	3
Damper Force	0.96	0.7	4	0.87	1.1	5
Tyre pressure	0.97	0.6	6	0.87	1.1	4
All variables expect for Grms and speed	0.94	0.8	-	0.82	1.4	-
D-class Sedan (DSD)						
Eliminated variable	Neural Networks		Relative Importance	Decision regression tree		Relative Importance
	R-squared	RMSE (m/km)		R-squared	RMSE (m/km)	
None	0.95	0.6	-	0.96	0.7	-

Grms	0.00	3.5	1	-0.26	3.8	1
Speed	0.87	1.3	2	0.74	1.7	2
N of People	0.96	0.7	4	0.95	0.8	5
Stiffness	0.96	0.8	3	0.93	0.9	3
DamperForce	0.97	0.6	5	0.95	0.8	4
Tyre pressure	0.97	0.6	6	0.95	0.7	6
All variables except for Grms and speed	0.92	0.9	-	0.91	1.1	-
Ignoring Vehicle Type						
Eliminated variable	Neural Networks		Relative Importance	Decision regression tree		Relative Importance
	R-squared	RMSE (m/km)		R-squared	RMSE (m/km)	
None	0.85	1.3	-	0.77	1.7	-
Grms	-0.02	3.5	1	-0.09	3.5	1
Speed	0.68	1.9	2	0.47	2.4	2
N of People	0.84	1.4	5	0.76	1.7	6
Stiffness	0.79	1.5	3	0.68	1.9	3
DamperForce	0.82	1.4	4	0.73	1.7	4
Tyre pressure	0.86	1.2	6	0.74	1.7	5
All variables except for Grms and speed	0.80	1.5	-	0.71	1.9	-

As can be seen from **Table 6.15**, in general, the machine learning algorithms performed better than the multivariable linear regression algorithms. The neural network performed better than the decision regression tree for all three datasets, as indicated by higher R-squared values and lower RMSE values. The feature eliminate process used for analysis of the machine learning algorithms showed that Grms (self-evidently) and speed have the

greatest impacts on the prediction of IRI. Removing the speed variable caused the RMSE for the neural network to increase from 0.6 m/km to 1.9 m/km for the LEV dataset and the R-squared value fell from 0.97 to 0.7.

In comparison, the analysis suggests that four other variables only have relatively minor influence on the prediction of IRI, this corresponds with the results of regression analysis described above.

The performance of the two algorithms on the three datasets were compared and the results showed that both algorithms performed better on the LEV and DSD datasets than on the combined dataset when none of these variables are eliminated. Removing the vehicle type variable from combined datasets resulted in the RMSE increasing by between 76% and 117% for both algorithms. This suggests that vehicle type is an important variable and should be taken into account when predicting IRI. In addition, by removing all variables that are considered to have medium effects (N of people, Stiffness, Damper Force, and tyre pressure) the performance of both algorithms do not appear to be greatly reduced. For example, for the neural network the R^2 value only increased by 3.0 % (0.96 to 0.92) for DSD dataset and 3.4% (0.97 to 0.94) for LEV dataset.

Section 6.7 Discussion

This Chapter has addressed objective 3 of the research, and by so doing it summarised and quantified (i.e. via a literature review and through vehicle numerical simulation

respectively) a number of influencing factors affecting the performance of a smartphone-based road roughness measurement system. These factors were associated with the smartphone itself (i.e. Section 6.2), the vehicle (i.e. Section 6.3) and the shape of the road profile (i.e. Section 6.4).

The suitability of using a smartphone fitted inside a moving vehicle to determine road roughness (IRI) has been assessed in this research by means of a review of existing approaches allied to a parametric study carried out with the aid of a vehicle dynamics package. It was found that an important aspect of measuring road roughness from a smartphone is to convert the smartphone's vertical acceleration data to Root-Mean-Square body accelerations (i.e. Grms). The experimental results showed that since Grms was strongly correlated with the actual road IRI. In a real-world system the calculation of the Grms of millions of measured acceleration signals accruing from road users is more practical and less time consuming compared to alternative approaches suggested in the literature which require the transformation of the data to the frequency domain (e.g. PSD). The literature showed that the frequency of data collection specified by international standards for accelerometer based systems could be achieved by modern smartphones.

An analysis of the results from the parametric study showed that vehicle speed and type and Grms had a large effect ($> 30\%$) on the assessment of IRI. Sprung mass, suspension stiffness, longitudinal profile and tyre pressure a moderate effect ($>10\%$) whilst acceleration/braking, wheelbase length and suspension damping a relatively minor

influence (<10%).

Multivariate linear regression and machine learning algorithms were developed to predict IRI. The results suggest that a multivariate linear regression analysis of vertical acceleration data converted to Grms values, taking into account vehicle speed, could be used to predict IRI to an accuracy which is equivalent to that which might be expected by visual observation (i.e. a Class 4 device at IQL-4). Machine learning algorithms were found to provide a more accurate estimate of IRI than using multivariate linear regression. Incorporating a machine learning algorithm within a suitable smartphone-based system could enable road roughness to be measured in accordance with the requirements of a Class 3 device (see **Table 2.4**). For example, the neural network trialed was able to predict road roughness to a RMSE of between 0.8m/km and 0.9m/km when vehicle class and speed were considered by the algorithm, allowing IRI to be predicted to meet the requirements of IQL-3. Data at this IQL, and at the frequency of collection allowed for by a smartphone, would enable long term strategic (i.e. network level) road management decision making of primary road networks. This includes (M. W. Sayers *et al.*, 1986):

- The provision of a summary of the condition of an entire road network on a regular basis (e.g. annually)
- Data to be used within network level models that compare maintenance policies and assess road use and road agency costs. For example, IRI values at IQL-3 could be utilised within a decision support tool, such as the World Bank's standard for

road investment appraisal, HDM-4, to enable for example road maintenance strategies to be compared (H. G. Kerali, Odoki and Stannard, 2006). This aspect is explored further in **Appendix B**.

- Prioritize maintenance and rehabilitation schemes.

Relatively low cost data collection systems could therefore be developed to support strategic road asset management, which incorporate smartphones with similar characteristics inside a fleet of vehicles of similar types, travelling at normal traffic speeds, to routinely inspect the condition of a road network.

Without knowing the vehicle type and speed, Grms data could be analysed using a machine learning based approach to enable IRI to be predicted to a similar accuracy as would be expected from a visual inspection. Such data could be suitable for the routine analysis of the condition of local road networks using a crowd sourcing type of approach (World Bank Group, 2016).

The above notwithstanding, although the algorithms presented for the analysis of vertical vehicle body acceleration demonstrated promising results for the artificial datasets developed for the research, clearly they would need further testing, calibration and refinement to the conditions at hand for their practical application. Chapter six further demonstrated the how to choose the most appropriate machine learning algorithm that can be used suitably for data modelling concerning a large-scale dataset in this research.

Summary

This chapter has found that:

- Converting vehicle body acceleration to Grms greatly facilitates the prediction of road roughness.
- Vehicle body acceleration, smartphone type, vehicle type and speed were found to be the dominant factors that influence road roughness measurement.
- A data collection system which utilises smartphones fitted inside a moving vehicle to assess road roughness, can satisfy the frequency of data collection and accuracy requirements of a Class 3 providing vehicle speed is taken into account and a suitable data processing approach is utilised.
- Without knowing vehicle type and speed, Grms data could be utilised to determine road roughness to a similar degree of accuracy as could be achieved by visual inspection (i.e. IQL-4).
- To account for the effect of the shape of road profiles on the measurement of IRI using a smartphone, particularly for roads in good condition, the use of a filtering techniques should be considered.
- Machine based learning algorithms outperformed linear regressions techniques in predicting road roughness from the simulated dataset. Of the two machine learning algorithms trialled, the neural network performed better than the decision regression tree.

The findings of this chapter are used in the development of a computational model in which the dominant factors are considered in data modelling. This is described **Chapter 8**.

Chapter 7

SELECTING AN APPROPRIATE PREDICTIVE MODEL

Section 7.1 Introduction

This chapter demonstrates the process used in the research to choose an appropriate predictive model which can be used to determine road roughness from vehicle body vertical acceleration data. **Chapter 3** (see [Section 3.2.11](#)) found that the most suitable types of techniques for the task in hand are a group known as machine-learning algorithms. The literature review (**Chapter 3**) identified 12 machine-learning algorithms that potentially might be useful for predicting road roughness from a distribution of measured vehicle body accelerations, whereas **Chapter 6** has determined the factors associated with the collection of data using a smartphone which should be taken into account in the modelling process.

In order to implement a smartphone-based road roughness measurement that can precisely predict road roughness (i.e.; to IQL-1/2), a number of variables need to be accounted for, these include vehicle speed, vehicle type, and the number of passengers (see [Section 6.6](#)). However, it is likely in any smartphone based system that it may not be possible to account for these variables as it may not be possible to measure them, this is likely to reduce the accuracy in predicting precise values of IRI. However, a smartphone based system which measures road roughness to IQL-3/4 is still useful (see [Section 2.3.1](#)). Accordingly, the

performance of the algorithms was tested with respect to their ability to (i) provide precise measures of road roughness (i.e. by providing model parameters in a regression analysis) and (ii) to classify road roughness in terms of categories / ranges of values (i.e. very good, good, fair, poor and very poor).

The processes for selecting the most appropriate algorithm for regression and classification respectively are summarised below and in **Figure 7.1**, and described further in this chapter.

The processes required are:

1. Select by means of a literature review candidate machine-learning algorithms (see Section 3.2.11).
2. Construct the algorithms using a suitable computer software platform
3. Identify measures of performance which can be used to assess the selected algorithms
4. Evaluate the performance of the algorithms using a suitable data set.

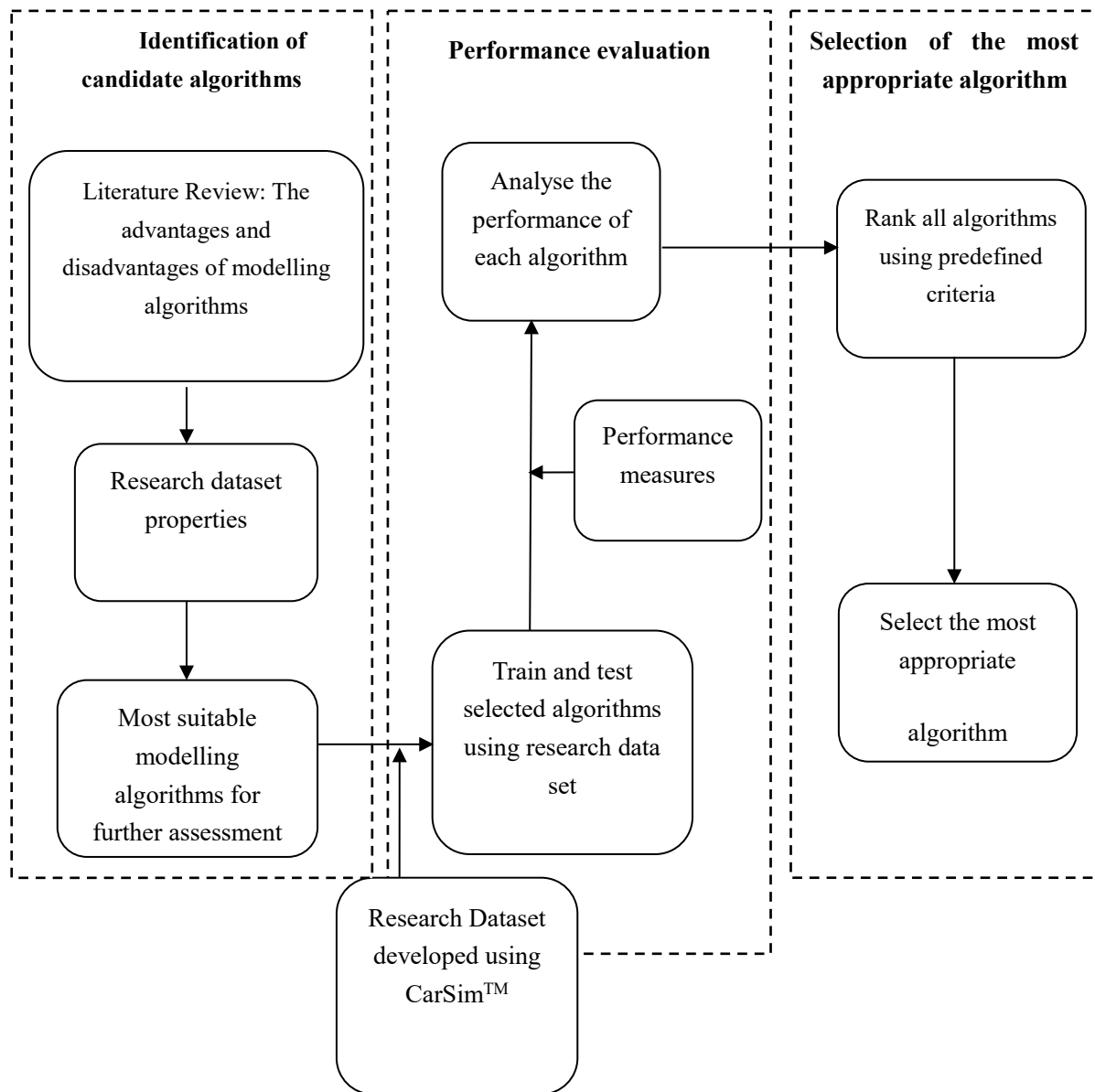


Figure 7.1: the selection process to choose the most appropriate algorithm

Section 7.2 Candidate Algorithms

The algorithms which were considered suitable for further consideration, with respect to the precise measurement of road roughness (i.e. for machine learning regression) and for classification are described in [Section 3.2.11](#) and listed below.

(i) For machine learning regression

- artificial neural network (ANN),
- random forests for regression,
- gradient-boosted regression tree,
- decision tree,
- linear and polynomial regression, and
- *K*-nearest neighbour (KNN).

(ii) For classification, the following machine-learning algorithms were trialled:

- Bayes naïve,
- random forest,
- neural network,
- support vector machine (SVM),
- gradient-boosted tree,
- decision tree, and
- K-Nearest Neighbour classification.

Section 7.3 Building the Algorithms

The KNIME Analytics platform was used to construct the candidate machine-learning algorithms (KNIME, 2018). The KNIME software is an open-source data analysis tool that provides a wide choice of advanced machine-learning algorithms. Its modular data-pipelining concept (i.e. a set of data processing elements connected in series, in which

the output of one element is the input to next one) enables users to assemble different components required for data processing using a graphical user interface. The KNIME platform is also powerful at processing large volumes of data (KNIME, 2018).

The values of important settings within the algorithms used for the trial are summarised in

Table 7.1.

Table 7.1: Specifications of important parameters in model setting

Regression problem		Classification problem	
Machine learning algorithms	Specification of important model parameters	Machine learning algorithm	Specification of important model parameters
Neural Networks	<ul style="list-style-type: none"> Maximum iterations =1000 Number of hidden layers = 2 Number of neurons per layer = 4 	Bayes Naïve	<ul style="list-style-type: none"> Default Probability (Laplace smoothing parameter) =0
Gradient-boosted Tree	<ul style="list-style-type: none"> Limited number of tree depth = 10 Number of model = 200 Learning rate = 0.1 	Random forest	<ul style="list-style-type: none"> Unlimited tree depth Number of model = 200
Random forest	<ul style="list-style-type: none"> Number of trees = 200 Unlimited tree depth 	Neural Networks	<ul style="list-style-type: none"> Maximum iterations =1000 Number of hidden layers = 2 Number of neurons per layer = 4
Polynomial regression	<ul style="list-style-type: none"> Maximum polynomial degree =10 	Support Vector Machine	<ul style="list-style-type: none"> Kernel: polynomial Parameters: power = 1.0; Bias = 1.0; Gamma = 1.0
Decision tree	<ul style="list-style-type: none"> Unlimited tree depth Unlimited node size 	Gradient-boosted Tree	<ul style="list-style-type: none"> Limited tree depth =10 Number of trees = 200 Learning rate = 0.1
K-nearest neighbour	<ul style="list-style-type: none"> K=10 	Decision tree	<ul style="list-style-type: none"> Unlimited tree depth Unlimited node size
		K-nearest neighbor	<ul style="list-style-type: none"> K=10

The performance of an algorithm can be affected by changing these parameters. For instance, increasing the number of iterations in neural networks, or the number of trees in the random forest method. Increasing the learning rate results in an increased number of trees (i.e. a higher learning rate generally requires more tree models to be analysed) in the gradient-boosted tree, which may also increase its performance. However, increased accuracy of an algorithm is generally associated with a longer training time. On the contrary, limiting the tree depth in the decision tree, random forest, and gradient-boosted tree or restricting the maximum iterations in neural networks can reduce the training time; however, these changes impact the accuracy of the algorithm.

Therefore, the values chosen for each of the parameters for each algorithm were determined on a trial and error basis so that a reasonable balance could be achieved between model accuracy and the length of the required training time.

Section 7.4 Performance Evaluation

To evaluate the performance of the candidate algorithms five criteria were developed as follows:

1. Accuracy of the model;
2. Ability to cope with large and complex datasets;
3. Training speed;
4. Generalisability of the algorithm (i.e. ability to avoid over-fitting)

5. Interpretability of developed model.

The rationale for including these five criteria in the model-selection process is as follows:

- An objective of the research is to identify a machine-learning algorithm which can predict road roughness, from a given dataset, to an appropriate degree of accuracy.
- The size of the dataset, typically at the network level, is considerable. Accordingly, the selected algorithm needs be able to cope with such dataset.
- The algorithm should be relatively quick to train and simple to use. The total time consumed in the processing of an extensive number of data to gain the desired output should be minimal, indicating good computational efficiency.
- The model needs to be transparent so that the user can have confidence in the results achieved by the model. That is, the model should be interpretable. Interpretability is the degree to which a human can consistently predict the result. A higher interpretability of a model indicates the easier that someone can comprehend why a model has made a given prediction
- A machine learning algorithm should be able to generalize from the training data to the entire domain of all unseen observations in the domain. An incapability of generalising, is known as overfitting of the training data (Cf. [Section 3.2.3](#)).

Using the above five criteria Equation 5.1 was developed to obtain a score P_k for each of the six proposed techniques with respect to their performance in determining roughness precisely and the seven techniques identified for classifying roughness values.

$$P_k = W_{accu} \times S_{accu} + W_{desg} \times S_{desg} + W_{speed} \times S_{speed} + W_{avoi} \times S_{avoi} + W_{intp} \times S_{intp} = \sum_{i=1}^n W_i \times S_{ik}$$

Equation 7.1

Where P_k is the weighted score of each algorithm calculated for rankings; W_{accu} , W_{desg} , W_{speed} , W_{avoi} , W_{intp} are the relative important of (weightings) *Accuracy*, *Design for large data*, *Training Speed*, *Avoid-overfitting* and *interpretability* five criteria, respectively, compared to each other, that is W_i , the relative importance of (weight) each performance measure i ; where S_{ik} is the score for a technique, k , according to the i th measure of performance. $\sum_{i=1}^n w_i = 1$, $1 \leq S_{ik} \leq 4$. The value of the subscript i relates to the numbered list of performance criteria given above (e.g. $i = 1$ refers to the accuracy of the model).

In order to determine the weights were deduced by pairwise comparison. The pairwise comparisons are transformed into a set of normalised weights by using the eigenvector procedure, for more details see (Saaty, 1990).

For all of the criteria considered, the score for a particular algorithm, P_k , was determined on a scale of 1 to 4. The higher the score the better the performance. For all of the performance measures except for accuracy, the score was based on the opinion of the researcher.

To assess accuracy the process described in Section 7.4.1 was adopted (Schlotjes, 2013).

Section 7.4.1 Measures of Accuracy

In machine learning, an algorithm's accuracy should not be evaluated by a single measure because the measure could be biased due to random success (Parker, 2011). Furthermore, a single evaluation of the data might show good accuracy by chance or by overfitting. To address these issues, first, a number of measures were selected to evaluate the accuracy of an algorithm. Second, a 10-fold cross validation process was applied in the evaluation process in which the training dataset was divided into 10 subsets and each subset was trained and tested in turn. The results of the 10 tests were averaged to present an overall score for each measure of accuracy.

For assessing algorithms used to infer a precise value of roughness (machine learning regression):

- coefficient of correlation (R^2),
- absolute error, and
- root mean square error (RMSE).

For the algorithms used to classify values into categories of roughness:

- absolute accuracy (percentage of correct classification),
- precision and recall (confusion matrix), and
- F -score.

The above measures are described below.

Machine Learning Regression

The Coefficient of determination (R^2 or R -squared)

The coefficient of determination measures how the variance in one variable can be explained by the variance in another variable within the regression analysis (i.e. how strong the linear relationship is between the two variables). It can be considered to be a measure of the likelihood of which future events can be predicted. The calculation of R -squared is defined as follows (Ferligoj and Kramberger, 1995):

$$R^2 = 1 - \frac{SS \text{ due to regression}}{SS \text{ about the Mean}} = 1 - \frac{\sum(\hat{y}_j - \bar{y})^2}{\sum(y_j - \bar{y})^2} \quad \text{Equation 7.2}$$

Where $y_j, j = 1, 2, 3 \dots n$, is the observed value of the dependent variable, and \hat{y}_j and \bar{y} denote the predicted value and mean, respectively. In addition, ‘SS about the mean’ is the sum of the squares of the deviation of the observed value from the mean, and ‘SS due to regression’ denotes the sum of squares of the deviation of the predicted value.

R -squared has a value of between 0 and 1 and becomes larger when a model achieves better fit (Ferligoj and Kramberger, 1995).

Mean absolute error and root mean square error

The mean absolute error (MAE) and root mean square error (RMSE) are commonly used error-based measures to determine the average performance of models (Chai and Draxler, 2014). MAE measures the average magnitude of error between predicted and observed

values without consideration of their direction (i.e. the absolute values of error), and the errors are averaged without consideration of the weight of each of them. In RMSE, the errors between predicted and observed values are initially squared and summed, and then the square root of the averaged summed errors is obtained. The calculations for MAE and RMSE are given in the equations below (Roy *et al.*, 2016):

$$MAE = \frac{\sum_{j=1}^N |p_j - a_j|}{N} \quad \text{Equation 7.3}$$

$$RMSE = \sqrt{\frac{\sum_{j=1}^N (p_j - a_j)^2}{N}} \quad \text{Equation 7.4}$$

Where p_j and a_j are the predicted Y value and observed Y value, respectively, for each observation, and N equals the number of observations in the test dataset.

For RMSE, squaring the errors will result in more weight being assigned to larger errors; however, MAE is a rather straightforward determination of prediction errors. Both MAE and RMSE are dimensioned statistics that range from 0 to infinity. They are defined as dimensioned statistics because they describe the average prediction differences between predicted and observed values in the units of the variable considered (Willmott and Matsuura, 2005). RMSE generally has a higher value than MAE. The greater the difference between them, the larger the variance of error in the prediction.

Classification

Confusion matrix

The confusion matrix is used to determine how well an algorithm can predict a classes (e.g. roughness is “good” or “fair”). **Table 7.2** shows the elements of a simple binary confusion matrix for an example of road condition classification.

Table 7.2: Confusion matrix (Han *et al.*, 2011)

True measurement			
Predicted Condition	Good		Fair
	Good	TP	FP
	Fair	FN	TN

The elements in the **Table 7.2** are defined as follows (e.g. for good condition):

- True positive (TP) of Good = all Good conditions that are correctly classified as Good;
- True negative (TN) of Good = all non-Good conditions that are not classified as Good;
- False negative (FN) of Good = all Good conditions that are not classified as Good; and
- False positive (FP) of Good = all non-Good conditions that are classified as Good.

The performance of measures of classification are calculated based on the confusion matrix that are shown as follows:

Absolute Accuracy

The absolute accuracy of a classifier for a given dataset measures the percentage of examples in the dataset that are correctly predicted by the classifier and represents the

performance of a classifier in general. The absolute accuracy is most effective; however, it is sometimes misleading, especially when each class in the dataset it is not uniformly distributed. The absolute accuracy is calculated as follows:

$$Accuracy = \frac{\text{examples are correctly predicted}}{\text{all examples in the testing data}} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{Equation 7.5}$$

Note that the error rate or misclassification rate of a classifier is calculated as *1-Accuracy*.

Precision and recall

Precision and recall are widely used in classification problems (Juba and Le, 2017). Precision can measure the percentage of predicted positive conditions that are true, and is referred to as *exactness* (e.g. the percentages of road profiles with conditions that are predicted as ‘good’ that are actually in good condition). A higher precision value indicates more confidence in the prediction (Parker, 2011; Sokolova and Lapalme, 2009).

Precision is determined as follows:

$$Precision = \frac{TP}{TP+FP} \quad \text{Equation 7.6}$$

Recall measures the percentage of road with true good condition that are classified correctly, and is also referred to as *completeness* (e.g. the percentage of roads with a true condition of ‘good’ that is also classified as ‘good’). This indicates the capability that a classifier can correctly identify the conditions of the road (Sokolova and Lapalme, 2009). A perfect score of 1 for the recall measure for the road condition ‘good’ indicates that all

road profiles with the true condition ‘good’ are correctly classified as ‘good’. However, the recall measure does not show how many other examples (e.g. ‘fair’ conditions) are incorrectly classified as ‘good’.

Recall is calculated as follows:

$$Recall = \frac{TP}{TP+FN} \quad \text{Equation 7.7}$$

F-score

An alternative to using both the precision and recall is to combine them into a integrated measure (i.e. the *F*-score), which is computed as follows:

$$F = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad \text{Equation 7.8}$$

The *F*-score is a weighted average of precision and recall, and is a preferred measure of performance over accuracy because it takes into consideration incorrect predictions (Sokolova and Lapalme, 2009). The closer the *F*-score is to 1 the more accurate the classifier.

K-fold cross validation

In supervised machine learning training data with known labels are used to develop a model that enables predictions based on data with unknown labels (testing data). Testing data is usually drawn separately from training data,

The algorithm searches for patterns/relationships between the features and targets in the training data and such patterns are used to make a new predictions based on testing data. However, the discovered patterns might not always be a true representation because the training data is unlikely to represent the whole population (Fernández *et al.*, 2015). A learning algorithm based on these spurious patterns usually performs better on the testing data than when it is applied to the whole population. This is known as overfitting (Elkan, 2013).

To address this cross validation is a commonly used approach to reduce overfitting of machine learning algorithms (Breiman, 2000). A *k-fold cross validation* is often used for this process and involves randomly partitioning the dataset into k subsets. The machine learning algorithm is trained using $k - 1$ subsets, while the one independent remaining subset is used for testing. This process is repeated k times until all subsets are tested. This process strongly increases the randomness (i.e. variance) in the data modelling which reduces a good accuracy by chance. The prediction error of each iteration is computed, and the final estimated performance is the average of the prediction errors over all k iterations.

Section 7.5 Datasets used for the trials

Actual road profiles obtained from the US State Highway Long-term Pavement Performance (LTPP) database (FHWA, 2018) were used as an input into CarSim™ to obtain a simulated dataset to trial and thereafter further develop the most appropriate machine-learning algorithm. In total, 430 road sections of a total length of 68.8 km were

selected to represent roads ranging from very good to poor condition.

Section 6.5 summarised the predominant factors that influence the measured vehicle body acceleration, and therefore road roughness measurement using a smartphone. While ideally all of these factors with high and medium effects should be considered when developing the dataset for testing, however, only those factors that can be easily measured in practice and can therefore be considered in any commercial system were considered in this research. Thus, factors such as suspension stiffness and tyre pressure were excluded because they can vary on a daily basis and it would not therefore be a practicable measurement for the road user to provide. Also, it would not have been possible to simulate the effect of smartphone type within a CarSimTM analysis. To account for these factors within the analysis, more data were randomly generated (within a limited range of variation) to increase the variance in the simulated dataset to approximate real data. The variables considered to develop the dataset are shown in **Table 7.3**.

A total of 168,480 simulations were conducted by varying the vehicle types (15), travel speeds (7), and number of passengers in a vehicle (4) for each road profile (430). Each simulation varied the value of each factor (vehicle type, speed, and number of passengers) one at a time while keeping the values of all other factors constant.

Table 7.3: Summary of dataset

<i>Attributes</i>	<i>Median</i>	<i>Minimum</i>	<i>Mean</i>	<i>Maximum</i>	<i>Standard Deviation</i>
N of People	2	1	2	4	1.12
Speed (km/h)	50	20	50	80	19.99
Vehicle type	15 types of vehicles vary from Utility Truck (UT) to F-class Sedan (FSD) (cf. Table 6.7)				
IRI (m/km)	4.77	0.83	4.66	9.16	1.88
Smartphones type	More data are generated to increase the variance (about 60%) of Grms measured by a particular vehicle type on a road at certain speed using CarSim TM				

Section 7.5.1 Data pre-processing

In order to improve the success of the machine-learning algorithms, the output data from CarSimTM simulations were pre-processed before they were used to train an algorithm.

The following data-processing techniques were used:

1. **Missing value:** All missing values were checked. If missing values were found, that row of data should be removed before inputting the dataset for the machine-learning algorithm for training.
2. **Normalisation:** In some of the machine-learning algorithms, the independent variables required for input may differ in terms of data form. Typically, some algorithms cannot handle nominal variables, such as vehicle type, in the dataset and but others require the independent variables to assume a value of between (0, 1). To address this, two data normalisation techniques were used to transfer the form of

the dataset into an acceptable format that can be used for training a model. The two normalisation techniques used were ‘normalised by decimal scaling’, which normalises data by dividing the value by 100, and ‘category to number’, which categorises the nominal variables into numerical values. The normalisation techniques used are summarised in **Table 7.4**.

Table 7.4: Normalisation techniques used

<i>Machine learning algorithms for regression</i>	<i>Required Normalisation techniques</i>
Neural Networks	1. Category to Number (the numbers 1-15 were used to represent each of the 15 vehicle types, e.g. 1 was used to represent UT and 15 to represent FSD see Table 6.7 . 2. Normalised by Decimal Scaling (e.g. Divide speed by 100 to make it between 0 and 1)
K-nearest neighbour	Category to Number (1-15 for vehicle types)
Linear and Polynomial regression	Category to Number (1-15 for vehicle type)
<i>Machine learning algorithms for classification</i>	<i>Required Normalisation techniques</i>
Neural Networks	1. Category to Number (1-15 for vehicle types) 2. Normalised by Decimal Scaling
Support Vector Machine	Category to Number (1-15 for vehicle type)
K-nearest neighbour	Category to Number (1-15 for vehicle type)

Section 7.6 Results of the model selection process

The 13 machine-learning algorithms were constructed in the KNIME platform as described in Section 7.3. The dataset described in Section 7.5 was used to assess the performance of each algorithm from the point view of determining precise values of roughness and for

classification respectively.

As mentioned above a 10-fold cross-validation approach was used to provide improved accuracy in the predictions and to reduce the issue of overfitting (Han *et al.*, 2011). The process involved randomly partitioning the data into 10 subsets of equal size (i.e. every 16,848 rows for each subset), and testing subset in turn. This process was repeated 10 times and the performance measures of the 10-fold cross-validated were averaged to provide a representative estimate for the algorithm used. The weighted scoring system described in Section 7.4 (see *Equation 7.1*) were used to rank the performances of the algorithms for the regression and classification problems. The results of these analyses are given below.

Regression Problem

Section 7.6.1 Measures of all criteria

Table 7.5 summarises the average measures of accuracy and training speed (defined in Section 7.4.1) of all algorithms used for the purposes of determining a precise measure of roughness. As seen from the results, the KNN, decision regression tree, random forest regression, and gradient-boosted tree methods outperform the neural network and linear and polynomial regressions, as evidenced by their higher *R*-squared values and lower MAE and RMSE. Of these four algorithms, the decision tree and KNN algorithms appear to have the greatest average performances (i.e. *R*-squared = 0.944 and 0.95 with MAE of

0.312 and 0.301 m/km respectively). The decision tree is also shown to have a very fast training speed (a total of two minutes); however, the KNN approach requires a relatively long time to train (four hours).

Table 7.5: The performances of the algorithms tested to determine precise values of roughness

<i>Machine learning algorithms</i>	<i>R-Squared (averaged of 10 cross-validation samples)</i>	<i>Averaged Mean Absolute Error (m/km)</i>	<i>Averaged Root mean Squared Error (m/km)</i>	<i>Time per run and time for 10-fold cross validation</i>
Neural Network	0.781	0.67	0.88	10 minutes per run and 3 hours in total
Gradient-boosted Tree	0.925	0.363	0.495	4.5 minutes per run with 48 minutes in total.
Random Forest	0.934	0.359	0.489	1 minutes 40 seconds per run and with 12 minutes in total
Linear regression	0.492	1.038	1.345	20 seconds per run and 2.5 minutes in total
Polynomial regression	0.376	1.115	1.473	17 seconds per run and 3 minutes in total
Decision tree	0.944	0.312	0.446	20 seconds per run and 2 minutes in total
K nearest neighbour regression (KNN)	0.95	0.301	0.434	25 minutes per run and more than 4 hours in total

Although the decision tree and KNN regression algorithms outperformed the others according to the metrics presented in **Table 7.5**, it is important to understand the relevant issues that are associated with these two algorithms. For the decision tree, small prediction

errors are often associated with extreme model complexity (see [Section 3.2.3](#)), i.e. they have *poor generalisability*. For instance, an extremely complex decision tree structure can be built with more than several thousand tree nodes and, in this case, excellent accuracy may be achieved. However, the tree structure generated may not perform as well when a new training or testing dataset is used (e.g. collected from the same population with a different distribution), i.e. there is an issue of *overfitting*. Moreover, another issue associated with extreme complex tree structure is reduced interpretability. Interpreting several thousand tree nodes to gain knowledge can be labour-intensive and impractical. Although cross validation can be used to reduce the risk of overfitting, typically, this algorithm still tends to over fit the training dataset in large-scale problems involving millions of data points (e.g. the smartphone-based road roughness measurement system) (Ignatov and Ignatov, 2017).

The KNN algorithm has natural characteristics such as lazy learning algorithms, in which all training data are kept in memory. New data are predicted based on their similarity (i.e. analogy) to the training data (Han *et al.*, 2011). This particular feature of KNN enables it to perform well in predicting IRI from a number of simulated Grms by considering important variables such as vehicle speed and types. However, the KNN has a number of disadvantages. First, the KNN was found to be computationally expensive (i.e. it runs slowly) when the sample size is large due to its distance-based comparison (Han *et al.*, 2011; Kotsiantis, Zaharakis and Pintelas, 2007). As can be seen from **Table 7.5**, a total

of 4 hours (25 minutes for each test) was required to process the 10-fold cross validation method to assess the model performance, whereas only 2 minutes were needed for the decision tree to perform the same tests. If the KNN is used to process very large datasets (e.g. several million), as is the case for the application proposed herein, it can be computationally expensive. This is why in the KNIME software the suitable range of data of analysis suggested for KNN is between a few thousand to maximum tens of thousands.

The random forest and gradient-boosted tree can address the issue of poor generalisability due to their inherent resampling functions, such as random selection or bootstrapping, as well as the powerful ‘majority vote’ and iterative function (Han *et al.*, 2011). These embedded features enable them to fit the large dataset well and be insensitive to noise. However, they suffer from poor interpretability (i.e. they are regarded as a black-box).

In the analysis, only the random forest and gradient-boosted tree algorithms can also cope with categorical features directly, i.e. the data-pre-processing (e.g. Normalisation for a number of ‘vehicle types’) processes are not required which reduces chance of the bias error caused by normalisation process (See **Table 7.4**).

As can be seen from **Table 7.5** the statistical regression (linear and polynomial) techniques generally are inferior to the other algorithms in terms of the performance and suitability for large datasets, as they do not cope with large variances, as evidenced by low R-Square and high RMSE scores. Furthermore, they were found to be sensitive to noise and changes in

the dataset. For example, in polynomial regression, better accuracy is normally associated with polynomials of a higher order, which can result in overfitting and noise. However, they also appear to have some advantages, such as a fast training speed and good interpretability (i.e. mathematical expression).

The neural network algorithms performed slightly better than the regression analysis in terms of measured R-square and RMSE. However, they do not show very good accuracy in testing compared to the random forest algorithm and are computationally expensive in processing large-scale datasets, such as the one used in this research.

Section 7.6.2 Weighted scoring system

As mentioned in Section 7.4 pairwise comparison was used to determine the relative importance of each measure of performance. Using this process the decision maker judges the relative importance of each measure by comparing it one by one with each other measure. Details of the pairwise comparison process are given in (Saaty, 1990). The results of the pairwise comparison (i.e. relative importance) for each of the five measures are shown in in **Table 7.6** and **Table 7.7**.

Table 7.6: Pairwise comparison matrix of the five criteria

A \ B	<i>Accuracy</i>	<i>Design for large Dataset</i>	<i>Training speed</i>	<i>Avoid-overfitting</i>	<i>Interpretability</i>
<i>Accuracy</i>	1.00	3.00	3.00	7.00	7.00
<i>Design for large dataset</i>	0.33	1.00	3.00	5.00	5.00

<i>Training speed</i>	0.33	0.33	1.00	3.00	3.00
<i>avoid-overfitting</i>	0.14	0.20	0.33	1.00	1.00
<i>Interpretability</i>	0.14	0.20	0.33	1.00	1.00
Sum columns	1.95	4.73	7.67	17.00	17.00

Note: “1” means criteria **A** is equally important than **B**

“3” means Criteria **A** is **moderately** more important than **B**, inversely, Criteria **B** is **slightly** less important than **A** (1/3).

“5” means **A** is **more** important than **B**, and inversely, Criteria **B** is **less** important than **A** (1/5).

“7” means criteria **A** is **strongly** more important than **B**, and inversely, Criteria **B** is **much** less important than **A** (1/7).

Accuracy was given the highest score as it was considered to be a standing point to select a good algorithm. The reason *design for a large dataset* was given a higher score than *training speed* is because any algorithm used must be able to process a large dataset (e.g. millions to hundreds of millions of data points) first, and then the one that has a faster training speed can be considered. *Interpretability* was given a larger score than *avoid overfitting* because algorithms with better generalisability (i.e. that avoid overfitting due to randomness in the algorithms) and performances are often associated with a complex model structure, causing issues of poor interpretability, particularly in processing a large dataset (e.g. random forest and gradient-boosted tree algorithms).

The above pairwise comparison matrix shown in **Table 7.6** was normalised by making the sum of each column to equal to 1, and the criteria weight was obtained by averaging each row. (see **Table 7.7**).

Table 7.7: Normalised pairwise comparison matrix for five criteria

	<i>Accuracy</i>	<i>Design for large Dataset</i>	<i>Trainin g speed</i>	<i>Avoid -overfitti ng</i>	<i>Interpretabili ty</i>	<i>Criteria Weights</i>	<i>Paramete rs</i>
<i>Accuracy</i>	0.51	0.63	0.39	0.41	0.41	0.47	W_{accu}
<i>Design for large dataset</i>	0.17	0.21	0.39	0.29	0.29	0.27	W_{desg}
<i>Training speed</i>	0.17	0.07	0.13	0.18	0.18	0.14	W_{Speed}
<i>avoid-overf itting</i>	0.07	0.04	0.04	0.06	0.06	0.06	W_{avoi}
<i>Interpretabi lity</i>	0.07	0.04	0.04	0.06	0.06	0.06	W_{intp}

Each algorithm was scored on a scale of 1 to 4 for each of five criteria according to the findings of Section 7.6.1 together with knowledge obtained from literature review (See **Table 7.8**).

Table 7.8: Scores of all algorithms in each measures

		S_{accu}	S_{desg}	S_{Speed}	S_{avoi}	S_{intp}
Rank	Representative scores	Accuracy	Design for large data	Training Speeds	Avoid over-fitting	Interpretabili ty
1 (best)	4	KNN, decision tree, Random forest and Gradient-bo osted tree	Random forest, Gradient-b oosted tree	Linear and polynomial regression, Decision tree and Random forest	linear regression, Random Forest and Gradient-bo osted tree	Linear and polynomial regression
2	3	Neural Network	Neural Network	Gradient-b oosted tree	Neural Network	Decision tree
3	2		Decision tree and KNN	Neural Network	Decision tree, KNN	KNN

4 (wors t)	1	Linear regression and Polynomial regression	Polynomial and linear regression	KNN	Polynomial regression	Random Forest, Gradient-boo sted tree and Neural Network
---------------------------------------	---	---	--	-----	--------------------------	---

The weights shown in **Table 7.7** together with the scores given in **Table 7.8** were combined using Equation 7.1, The resulting weighted scores for each algorithm according to the use of the algorithm for the precise measurement of roughness are show in **Table 7.9**.

Table 7.9: The weighted scores for all algorithms used to obtain precise values of regression

Weights	0.47	0.27	0.14	0.06	0.06		
<i>Technique</i>	<i>Accuracy</i>	<i>Design for Large Data</i>	<i>Training Speed</i>	<i>Interpretab ility</i>	<i>Avoid over-fitting</i>	<i>Score</i>	<i>Rank ing</i>
<i>NN</i>	3	3	2	1	3	2.74	5
<i>Random Forest</i>	4	4	4	1	4	3.83	1
<i>Decision tree</i>	4	2	4	3	2	3.29	3
<i>Gradient Boosting tree</i>	4	4	3	1	4	3.69	2
<i>KNN</i>	4	2	1	2	2	2.80	4
<i>Linear analysis</i>	1	1	4	4	4	1.77	6
<i>Polynomial regression</i>	1	1	4	4	1	1.60	7

It may be seen from **Table 7.9** that the random forest regression algorithm outperformed the others, as evidenced by the highest score (i.e. 3.83). The random forest algorithm has advantages including good accuracy and fast training speed (see **Table 7.5**); it eases in setting parameters and randomisation in dealing with a large dataset (i.e. avoiding overfitting), although it has one perceived disadvantage (poor interpretability).

The gradient-boosted tree approach was also found to offer advantages and the evaluation of its performance resulted in a similar overall score to the random forest approach (i.e. score of 3.69). However, it is more computationally expensive than the random forest algorithm, requires a longer training time (see **Table 7.5**) and needs more parameters to be specified (e.g. learning rate, number of trees, and tree depths).

Accordingly, the random forest algorithm was selected as the most appropriate machine-learning algorithm for obtaining precise values of road roughness

However, it could be argued that the ranking of machine-learning algorithms is dependent on the weights assigned to the five criteria and that changing the weights might affect the rankings. To address this question, the weights of five criteria were reallocated. This is further elaborated in the discussion (see [Section 7.7](#)).

Classification problem

Section 7.6.3 Measures of Accuracy and Training speed

The score for accuracy and training speed of available machine-learning classifiers according to the criteria provided in Section 7.4.1 are summarised in **Table 7.10**.

Table 7.10: measures of classification algorithms

Machine learning algorithms	Accuracy (%)	Precision (Averaged)	Recall (Averaged)	F-score (averaged)	Time consumed for analysing
<i>Bayes naive</i>	49.9	0.253	0.291	0.270	10 seconds per run and 2 minutes in total
<i>Random Forest</i>	90.1	0.902	0.887	0.893	2 minutes per run and 25 minutes in total
<i>Neural Network</i>	65.8	0.709	0.642	0.665	1 minutes and 11 minutes in total
<i>SVM</i>	-	-	-	-	Extremely long (minimum of 4 hours)
<i>Gradient-boosted Tree</i>	88.3	0.865	0.871	0.868	5 minutes per run and 52 minutes in total
<i>Decision tree</i>	90.0	0.899	0.891	0.894	15 seconds per run and 3 minutes in total
<i>K nearest neighbour Classification</i>	91.0	0.917	0.900	0.907	10 seconds per run and 2 minutes in total

As can be seen from **Table 7.10**, the performances of the random forest, gradient-boosted tree, decision tree, and KNN algorithms were the best, as reflected by their higher accuracy (i.e. percentage of correct predictions) and higher *F*-score (i.e. the weighted average of the precision and recall). The neural network and naïve Bayes algorithms showed poorer results than the four other classifiers. The accuracy and *F*-score were 65.8 % and 0.656, respectively, for the neural network and only 49.9 % and 0.27, respectively, for the naïve Bayes algorithm.

Among the four most promising classifiers, KNN had the highest accuracy (i.e. 91%) and *F*-score (0.907), and the gradient-boosted tree has the lowest accuracy (i.e. 88.2%) and *F*-score (0.868). The decision tree and KNN approaches showed the fastest training speeds of 3 minutes and 2 minutes, respectively. However the random forest required a longer training time (i.e. 11 minutes), whereas the gradient-boosted tree was found to be slower in training speed than the random forest algorithm (i.e. 52 minutes to train).

The SVM classifier was intended to be included to compare with other classifiers; however, it took an extremely long time (i.e. at least 4 hours) to train and therefore was not considered further.

Section 7.6.4 Other measures of performance

The results obtained in Section 7.6.3 for algorithm accuracy and training speed and those for the other three measures of performance were rebased using a scale of 1 to 4 for each

algorithm (See **Table 7.11**). The resulting weighted scores (using Equation 7.1) for each algorithm taking into account their overall performance are shown in **Table 7.12**.

Table 7.11: Rankings of classification algorithms

<i>Rank</i>	<i>Representative score</i>	<i>Accuracy</i>	<i>Design for large data</i>	<i>Training Speeds</i>	<i>Avoid over-fitting</i>	<i>Interpretability</i>
Level 1 (best)	4	KNN, decision tree, Random forest and Gradient-boosted tree	Random forest, Gradient-boosted tree	Naïve Bayes, KNN, Decision tree and Neural networks	Naïve Bayes, Random Forest and Gradient-boosted tree	
Level 2	3	Neural network	Neural Network	Random Forest		Decision tree
Level 3	2		Decision tree and KNN	Gradient-boosted tree	Neural Network	
Level 4 (worst)	1	Naïve Bayes	Naïve Bayes	SVM	Decision tree, KNN	Random Forest, Gradient-boosted tree, Neural Network and Naïve Bayes

Table 7.12: Weighted scores for classification algorithms

<i>Weightings</i>	<i>0.47</i>	<i>0.27</i>	<i>0.14</i>	<i>0.06</i>	<i>0.06</i>		
<i>Technique</i>	<i>Performan ce</i>	<i>Design for Large Data</i>	<i>Training Speed</i>	<i>Avoid over-fitting</i>	<i>Interpret ability</i>	<i>Score</i>	<i>Ranking</i>
<i>NN</i>	3	3	4	2	1	2.96	5
<i>Random Forest</i>	4	4	3	4	1	3.68	1
<i>Decision tree</i>	4	2	4	1	3	3.22	3
<i>Gradient Boosting tree</i>	4	4	2	4	1	3.54	2
<i>KNN</i>	4	2	4	1	2	3.16	4
<i>Naïve Bayes</i>	1	2	4	4	1	1.87	6
<i>SVM</i>	–	–	–	–	–	–	

From the results in **Table 7.12** it is evident that the random forest algorithm is the best performing classifier of all algorithms considered due in part to its great ability in coping with errors and noise in a large dataset. The embedded ‘majority vote’ function and the random resampling of the training data process in the random forest algorithm enables it to deal with large-scale database, as would be needed for the prediction of road roughness determined from on vehicle body acceleration measured by the number of road users using smartphones. Therefore, it was selected as the most appropriate classifier to be used to build the prototype model.

Section 7.7 Discussion

In this chapter, a model-selection process was conducted via a comparative study of different machine-learning algorithms. The results showed that the random forest algorithm performed the best of those considered in predicting road roughness values precisely (i.e. for IQL-1/2) and as a classifier (i.e. IQL-3/4). It could be argued however that the weights of the five criteria might affect the performance and therefore the ranking of the algorithms. To address this, an analysis was carried out of the potential impact of the weights on the performance of the six selected algorithms for the precise measurement of road roughness. Two different groups of weights for the five criteria were considered as shown in the **Table 7.13** and **Table 7.14**.

Table 7.13: Pairwise comparison matrix of the five criteria in first group

A \ B	B	<i>Accuracy</i>	<i>Design for large Dataset</i>	<i>Training speed</i>	<i>Avoid-overfitting</i>	<i>Interpretability</i>
<i>Accuracy</i>		1.00	1.00	1.00	7.00	7.00
<i>Design for large dataset</i>		1.00	1.00	1.00	7.00	7.00
<i>Training speed</i>		1.00	1.00	1.00	7.00	7.00
<i>avoid-overfitting</i>		0.14	0.14	0.14	1.00	1.00
<i>Interpretability</i>		0.14	0.14	0.14	1.00	1.00
Sum columns		3.29	3.29	3.29	23.00	23.00

Note: “1” means criteria **A** is equally important than **B**

“3” means Criteria **A** is **moderately** more important than **B**, inversely, Criteria **B** is **slightly** less important than **A** (1/3).

“5” means **A** is **more** important than **B**, and inversely, Criteria **B** is **less** important than **A** (1/5).

“7” means criteria **A** is **strongly** more important than **B**, and inversely, Criteria **B** is **much** less important than **A** (1/7).

Table 7.14: Pairwise comparison matrix of the five criteria in the second group

A \ B	B	<i>Accuracy</i>	<i>Design for large Dataset</i>	<i>Training speed</i>	<i>Avoid-over fitting</i>	<i>Interpretability</i>
<i>Accuracy</i>		1.00	3.00	3.00	7.00	5.00
<i>Design for large dataset</i>		0.33	1.00	1.00	5.00	3.00
<i>Training speed</i>		0.33	1.00	1.00	5.00	3.00
<i>avoid-overfitting</i>		0.14	0.20	0.20	1.00	0.33
<i>Interpretability</i>		0.20	0.33	0.33	3.00	1.00
Sum columns		2.01	5.53	5.53	21.00	12.33

Note: “1” means criteria **A** is equally important than **B**

“3” means Criteria **A** is **moderately** more important than **B**, inversely, Criteria **B** is **slightly** less important than **A** (1/3).

“5” means **A** is **more** important than **B**, and inversely, Criteria **B** is **less** important than **A** (1/5).

“7” means criteria **A** is **strongly** more important than **B**, and inversely, Criteria **B** is **much** less important than **A** (1/7).

After normalising the values in the two matrixes (see above), two news groups of criteria weights were established as shown in **Table 7.15**. For comparison, the weights used for the regression problem in section 7.6.2 are also given.

Table 7.15: Two groups of weights used

Criteria	Parameters	Original weights used in regression	First group of weights	Second group of weights
Performance	W_{accu}	0.47	0.30	0.46
Design for large data	W_{desg}	0.27	0.30	0.20
Training speed	W_{speed}	0.14	0.30	0.20
Interpretability	W_{avoi}	0.06	0.05	0.09
Avoid overfitting	W_{intp}	0.06	0.05	0.04

The reason these two groups of weights were chosen is because they are representative scenarios which may occur in practice. One is termed here in as optimistic and the other could be considered a pessimistic scenario. In the optimistic scenario, it is assumed that a large dataset is available. Therefore the corresponding group of weights are based on the assumption that the accuracy, design for large data, and training speed are equally important when implementing a smartphone-based road roughness measurement system. However, the remaining two criteria are considered be less important because of the existence of the large accurate data set. The data in such a huge dataset are assumed to be capable of being updated at any time interval and can be used as a training dataset. Therefore, the problems of overfitting and less interpretability are of relatively less concern because ever-changing distribution of the training dataset reduces the risk of overfitting, and also, it can become more difficult to interpret when the training dataset continually changes.

The results of the weighted scoring system using the first group of weights for the five criteria are shown in **Table 7.16**.

Table 7.16: Weighted scoring for all regression algorithms

Weightings	0.3	0.3	0.3	0.05	0.05		
<i>Technique</i>	<i>Accuracy</i>	<i>Design for Large Data</i>	<i>Training Speed</i>	<i>Interpretability</i>	<i>Avoid over-fitting</i>	<i>Score</i>	<i>Rank</i>
<i>NN</i>	3	3	2	1	3	2.61	4
<i>Random Forest</i>	4	4	4	1	4	3.87	1
<i>Decision tree</i>	4	2	4	3	2	3.26	3
<i>Gradient Boosting tree</i>	4	4	3	1	4	3.57	2
<i>KNN</i>	4	2	1	2	2	2.30	5
<i>Linear analysis</i>	1	1	4	4	4	2.17	6
<i>Polynomial regression</i>	1	1	4	4	1	2.04	7

Despite changing the weights, **Table 7.16** shows that the random forest algorithm still outperforms the others. The gap in the scores between random forest and gradient-boosted tree algorithms has noticeably increased compared the original analysis presented in **Table 7.9**.

When determining the second group of weights (i.e. the pessimistic scenario), most emphasis was put on the accuracy of prediction where the weight of interpretability was

considered to be greater than the criteria *avoid-overfitting* because the model needs to be transparent. The criteria *design for large data* and *training speed* were assigned weights smaller than *accuracy* (see **Table 7.14**). It is not suggested to assign *interpretability* the same weight as *training speed* and *design for large data*. In fact, the interpretability of the machine-learning algorithm in dealing with such a large amount of data (i.e. millions to hundreds of millions of data points) becomes less attractive to users because the training dataset is updated and changed frequently. The results of the weighted scoring system using the second group of weights are shown below.

Table 7.17: Weighted scoring for all regression algorithm

Weightings	0.46	0.2	0.2	0.09	0.04		
<i>Technique</i>	<i>Accuracy</i>	<i>Design for Large Data</i>	<i>Training Speed</i>	<i>Interpretability</i>	<i>Avoid over-fitting</i>	<i>Score</i>	<i>Ranking</i>
<i>NN</i>	3	3	2	1	3	2.62	5
<i>Random Forest</i>	4	4	4	1	4	3.73	1
<i>Decision tree</i>	4	2	4	3	2	3.42	3
<i>Gradient Boosting tree</i>	4	4	3	1	4	3.53	2
<i>KNN</i>	4	2	1	2	2	2.73	4
<i>Linear analysis</i>	1	1	4	4	4	2.00	6
<i>Polynomial regression</i>	1	1	4	4	1	1.87	7

Table 7.17 illustrates that the random forest algorithm still outperforms the other approaches. However under this scenario, the score for the decision tree has increased due to its comparatively good interpretability. Accordingly the results show that for both scenarios the random forest algorithm is still the most appropriate algorithm of those considered for the task in hand.

Although the random forest trained faster than Gradient-boosted tree (GBT), the performance of GBT was found to be very similar to the random forest algorithm (see **Table 7.9**). This is because the gradient-boosted tree algorithm is similar in the way it functions to the random forest algorithm. Both are ‘tree ensemble’ learners. The random forest algorithm builds trees in parallel with the majority vote function that was used in predictions, whereas the gradient-boosted tree creates a series of ‘weak’ trees, and the performance of the predictions is gradually improved by adding more trees (Olinsky *et al.*, 2014).

However, the performance of GBT is sensitive to the parameters in the model settings, The important parameters required to construct the gradient-boosted tree algorithm includes tree depths limits, number of models and learning rate, whereas for the random forest algorithm, the only parameter is the number of trees (models).

A second question to consider is whether it is possible that the training speed of the random forest algorithm can be improved without reducing the overall performance and

overtake the random forest. To address this, an analysis was undertaken by changing the number of trees in the RF algorithm. The results are shown in **Table 7.18**.

Table 7.18: *Accuracy of random forest algorithm under different model settings*

Random Forest			
Number of models	Training time per run	R-squared	MAE
200	2 minutes 30 seconds	0.935	0.351
100	1 minutes 20 seconds	0.934	0.353
50	40 seconds	0.932	0.357

From **Table 7.18** it can be seen that reducing the number of models in a RF could result in a large reduction in training time without adversely affecting the accuracy of the algorithm. This suggests that careful selection of the number of models in an RF algorithm can help to optimize the overall performance of the algorithm. In the next chapter, a sensitivity analysis is carried out to determine the appropriate number of models.

Section 7.8 Summary

This chapter has demonstrated a model-selection process to choose appropriate machine-learning algorithms via a comparative study among different machine-learning algorithms. The comparative study was undertaken by building a weighted scoring system to rank the integrated performances of different machine-learning algorithms developed to (i) compute the precise determination of roughness (i.e. machine learning based regression)

and, (ii) classify data into bands of roughness. Six candidate algorithms were tested on a trail dataset developed using CarSimTM for the former and seven for the latter. Five criteria were used for this process, namely: accuracy, design for large data, training speed, interpretability, and avoiding overfitting.

According to the results from the weighted scoring system, the random forest algorithm was identified as the most appropriate machine-learning algorithm for both the regression and classification problems. The use of the random forest algorithm for the task at hand is described in Chapter Eight.

Chapter 8

THE DEVELOPMENT OF PROTOTYPE SYSTEM

Section 8.1 Introduction

This purpose of this research is to develop a prototype system that can utilize data captured using smart phone technology to determine at suitable degrees of accuracy, repeatability reproducibly and information quality level (IQL), the appropriate measures of road condition, in order to support decision making in road asset management. Chapter four described a framework for the development of this smartphone-based prototype system, while Chapters six and seven presented the required components of the system. The goal of this Chapter and the subsequent one, is therefore to describe the development of a comprehensive system that combines and utilizes the findings inform Chapters six and seven to address the research objective four. (i.e. to build a prototype system that accurately and repeatedly determines an appropriate measure of road roughness from data captured by road users using smartphone technology, in which the predicted road roughness information can be used to support decision making for long term strategic road management). This chapter focuses on the Random Forest computational model selected in Chapter seven as the most appropriate of the methods identified from the comparative analysis for predicting road roughness from measured vehicle body acceleration. In particular, this chapter describes:

- 1) Obtaining a simulated dataset to trial and further develop the Random Forest algorithm, including:
 - (a) Selecting the variables to be included in the simulated dataset.
 - (b) Carrying out a number of simulations using CarSim™ to obtain the dataset
- 2) Building the Random Forest algorithm.
- 3) Evaluating the overall performance of the selected Random Forest algorithm which included:
 - (a) Choosing the appropriate information quality level to achieve the appropriate level of data accuracy
 - (b) Carrying out a number of statistical analyses to compare the IRI values predicted by the Random Forest algorithm with actual IRI values.

Section 8.2 Datasets for developing the computational model

Section 8.2.1 The Variables Considered

The variables considered in the simulated dataset were vehicle type and speed, vehicle sprung mass (i.e. to reflect the number of people in, or load of, the vehicle) and the Grms of vehicle body acceleration. Section 6.6 analysed the effect of a number of factors (variables) which can affect the measurement of Grms and therefore of road roughness (see *Table 6.12*). Whilst all of these factors could have been incorporated into the analysis of the effectiveness of the selected Random Forest algorithm it was decided to choose only

those factors which in practice can be easily measured and therefore taken into account in any commercial system. Two examples of such factors which were excluded are suspension stiffness and tyre pressure, values of which could vary on a daily basis and which would not be practicable therefore for the road user to provide.

Although this process simplifies the data acquisition process as the data generation is extremely time consuming, it puts more emphasis on the Random Forest algorithm to predict road roughness accurately using fewer variables. The factors included in the study, and those excluded, are shown in **Table 8.1**. All of those factors which were found to have a major effect on vehicle Grms in Section 6.6 are included.

In addition to the above, to take into account the effects of different smartphone types which were found in **Chapter 6** to have a large effect on the prediction of roughness at the large network-level data acquisition, in the CarSimTM simulations, different distributions of errors were generated to increase data variance of the Grms measured by a particular vehicle type on a road at certain speed. By so doing this allowed the potential influence on the measurement of vertical acceleration caused by the variations of the type of smartphone to be taken into account (cf. Section 8.3.1.3). This would not have been possible to simulate within a CarSimTM analysis. Another reason that the suspension stiffness and tyre pressure are not directly considered as variables in the CarSimTM data simulation is because they are found to have relatively small effect on the predicted road roughness (See Section 6.6), and also, these factors requested special vehicle mechanic

tool for measurements which would be impractical to measure at network-level.

Table 8.1: the variables included and excluded from the simulated dataset

Variables included	Variables excluded
Vehicle speed and type, sprung mass, Grms; and the effect of smartphone type	Variables with a medium or small influence on Grms including suspension damping, stiffness, tyre pressure, accelerating/braking, wheelbase and placement of smartphone

Section 8.2.2 Road Profiles

Road profile data were gathered from the United State Highway LTPP (FHWA, 2018) road database. In total 2, 059 road sections of a total length of 330km were obtained from the LTPP database to represent roads ranging from very good to poor condition. The IRI values of the roads were determined from their profiles using a computer program written by the researcher in Visual Basic for Applications (VBA). The computer program is based on guideline provided for determining IRI from road profiles, for more details see (M. W. Sayers, 1995).

Section 8.2.3 CarSim™ Simulation

A dataset that relates IRI to simulated Grms as a function of the influencing factors selected above was obtained using CarSim™. The road profiles obtained from the

State Highway LTPP database were input into CarSimTM and a number of simulations were carried out using different vehicle types, travelling at various speeds and with differing numbers of passengers for each road profile. Each simulation consisted of varying one value of each factor (i.e. car type, speed, number of passengers) at a time whilst keeping the values of all other factors constant. This resulted in a total of 807, 344 sets of data. The ranges of values of the variables used are shown in **Table 8.2**, together with some associated descriptive statistics.

Table 8.2: Summary of the datasets

Attributes	Median	Minimum	Mean	Maximum	Standard Deviation
N of People	2	1	2	4	1.12
Speed (km/h)	50	20	50	80	19.99
Vehicle type	15 types of vehicles varying from Utility Truck (UT) to F-class Sedan (FSD) as summarised in Table 6.7				
IRI (m/km)	4.77	0.83	4.66	9.16	1.88
Smartphones type	More data are generated to increase the variance (about 60%) of Grms measured by a particular vehicle type on a road at certain speed using CarSim TM				

Figure 8.1 shows the distribution of road sections obtained in terms of an IRI frequency plot. The information is also summarised in **Table 8.3**. Using the rating system suggested by RNET (Archondo-Callao, 2009) and given in **Table 8.4**, the overall condition of the simulated road sections is summarised in **Figure 8.2**. From **Figure 8.2**, it may be seen that 7.5% of the simulated road sections are in very good condition and 20%

are in good condition, approximately 48% are in fair condition and 25% are in poor condition.

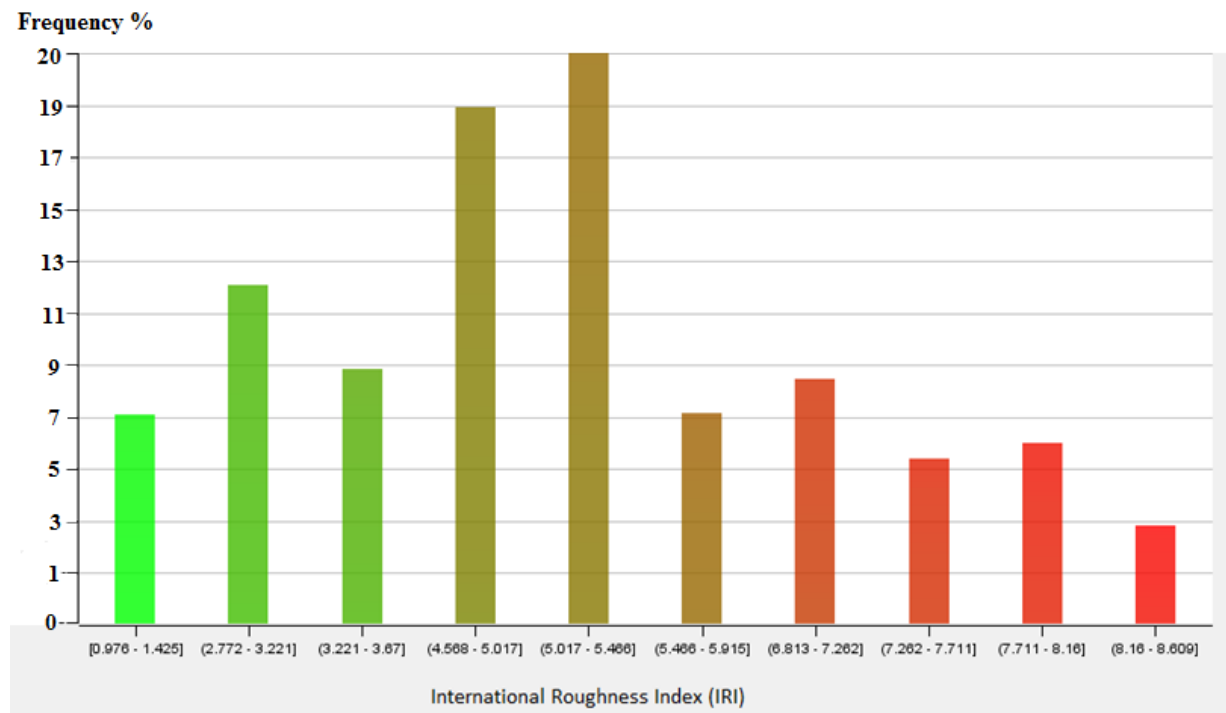


Figure 8.1: Distribution of datasets in the research

Table 8.3: Road IRI Classification (Archondo-Callao, 2009)

Condition category	IRI ranges (m/km)	Definitions (paved roads)
Very Good	0 - 2.5	Roads in very good condition with no capital road works.
Good	2.5 - 3.5	Roads are largely free of defects, requesting some minor maintenance activities, such as crack sealing.
Fair	3.5 - 6	Roads with defects and weak structural resistance, requesting resurfacing but no need for reconstruction.
Poor	6 – 10	Roads request rehabilitation or partial reconstruction.
Very Poor	10 - 16	Roads request full reconstruction, nearly equivalent to new construction.

Table 8.4: *A summary of dataset in terms of road condition*

Road Condition	IRI (m/km)	Mode (m/km)	Number of Data	Number of sections (160m)	Length(km)
Very Good	0-2.5	1.01	60490	154	24.7
Good	2.5-3.5	3.14	161164	411	65.8
Fair	3.5-6	4.97	387164	988	158
Poor	6-10	7.17	198526	506	81
Very Poor	10-16	-	-	-	-
In total		-	807344	2059	330

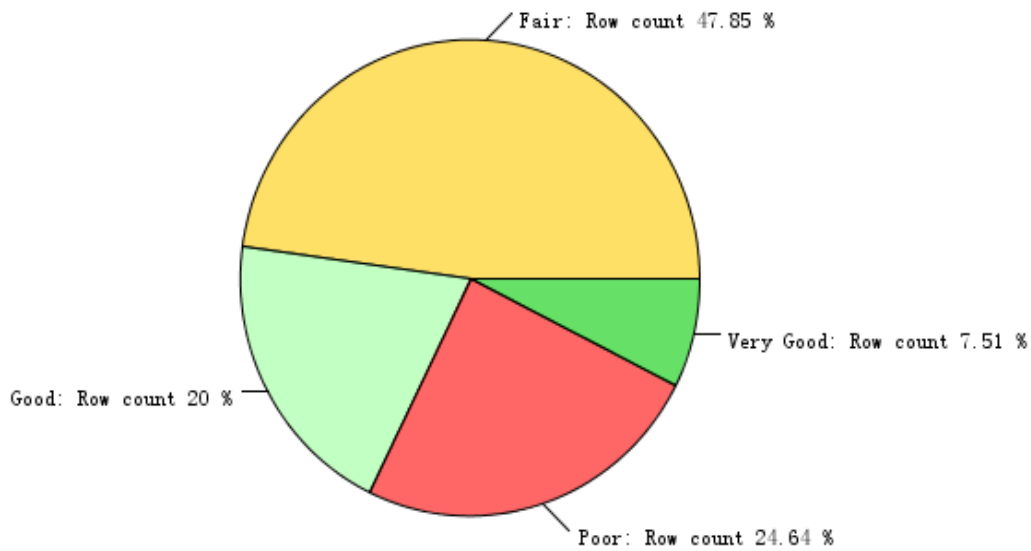


Figure 8.2: *Overall road conditions of network summarised in the research dataset*

Fifteen vehicle types were used to represent the majority of vehicle types in the UK market (see **Table 6.7**). In the CarSimTM data simulations, all vehicles of different vehicle type were uniformly distributed (approximately 7% each vehicle type and 100% in total), therefore the amount of vehicle under each vehicle type is very similar to eliminate the

bias prediction error caused by unbalanced traffic. This is because traffic composition (i.e. the distribution of different vehicle types in percentage terms) may affect the accuracy of prediction of IRI, for example, more commercial vehicles (e.g. Utility truck and European van) in the datasets tend to increase the error of prediction (cf. [Section 6.3.2](#)).

Relationship between independent and dependent variables:

The Pearson's linear correlation was used to indicate the linear relationship between the modelled inputs/independent variables and the outputs/dependent variables (G. Hall, 2015).

The Pearson's correlation is considered to be a measure of the linearity between inputs and outputs where a value of the index of -1 or 1 indicates a perfect relationship (i.e. negative or positive relationship respectively). A value of zero indicates no relationship. The comparisons for the generated data are shown in the *Table 8.5*.

Table 8.5: The Pearson's correlation matrix between attributes and outputs

	Sprung mass	Speed	IRI	Vehicle type	Grms
Sprung mass	1	0	-0.0009	×	-0.095
Speed	0	1	0	×	0.455
IRI	-0.0009	0	1	×	0.543
Vehicle type	×	×	×	1	×

Grms	-0.095	0.455	0.543	×	1
corr ■ = -1 corr ■ = +1 corr × = n/a					

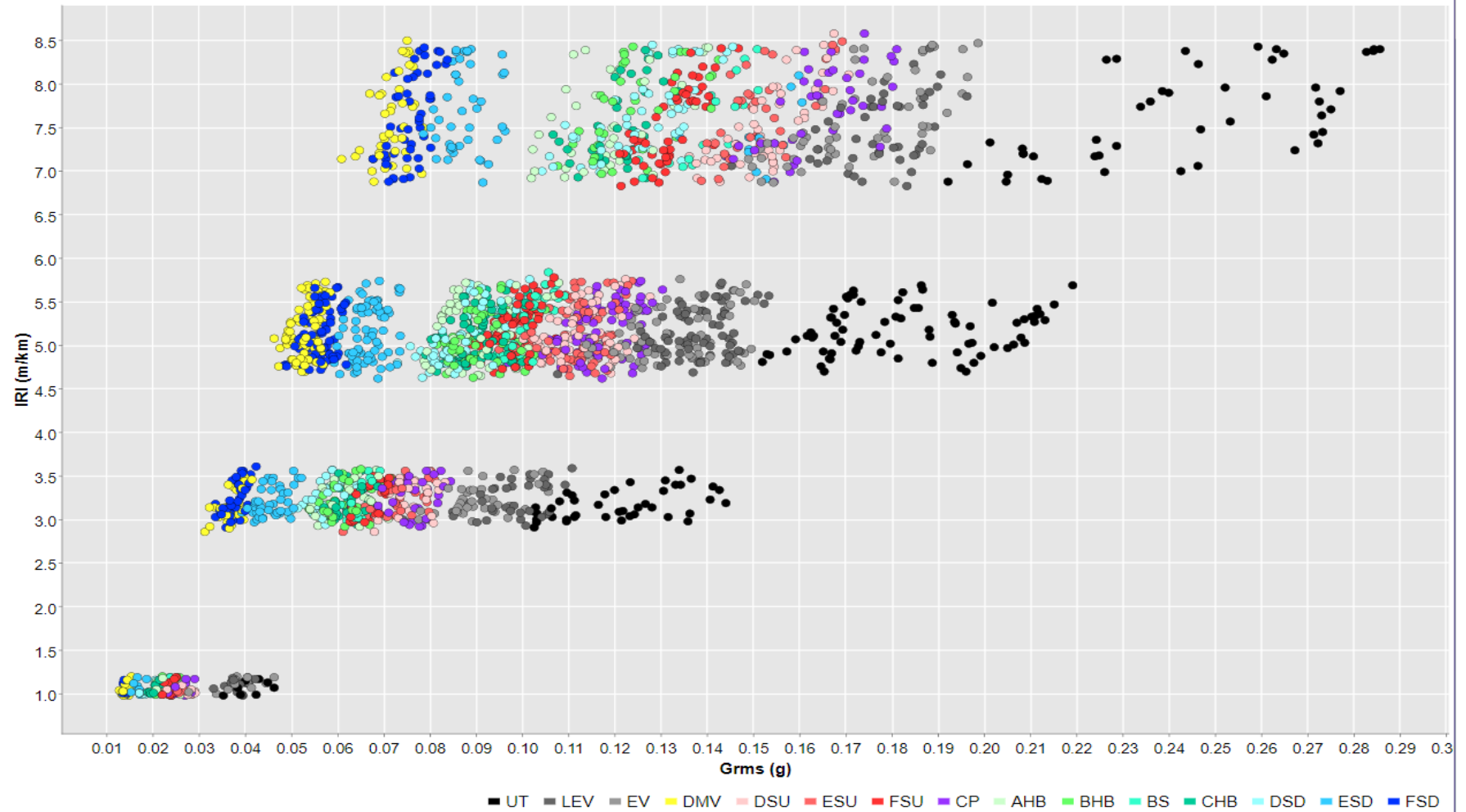
Considering the dataset as a whole, it may be seen that there is a positive correlation between Grms and IRI of 0.543. I.e. roads with higher IRI values cause higher vehicle vertical body accelerations, as would be expected. Speed and Grms are also positively related to each other (i.e. 0.455), suggesting that the faster a vehicle travels, all other factors being equal, the greater the vertical vehicle body acceleration. On the other hand, changes to the sprung mass (used to represent additional cargo / passengers) were found to be negatively correlated (about 0.1) with Grms. This may be expected as the heavier the vehicle, the greater its inertia and therefore the less the vertical acceleration caused by the road roughness (cf. [Section 6.3.5](#)).

Since vehicle types are nominal variables they cannot be reflected by a Pearson's correlation. Therefore to explore the relationship between vehicle type, IRI and Grms, a small portion of the data (2500 sets of data) were selected using stratified sampling together with random selection techniques to reduce bias due to sampling (Singh, 2003). The results are plotted in the form of a scatter matrix in **Figure 8.3**. For comparison, all data plotted were selected at for a vehicle speed of 50 km/h.

Figure 8.3: Scatter Plot of IRI against Grms with colours that represent different vehicle types

Scatter Plot (IRI vs Grms)

different color represents various vehicle types



In **Figure 8.3** the different colours represent different vehicle types. In general, the simulated Grms of all vehicles increases with IRI (as expected). The simulated Grms values of passenger cars (e.g. BS, CHB, AHB, ESD and etc) are clustered on the left hand side of the chart i.e. they show lower Grms values while the data for the heavier vehicles (e.g. Utility Truck, European Van, Large European Van, Compact Pickup) are grouped on the right hand side. This suggests that the simulated Grms values for the trucks and vans will be larger and vary in value more than those for passenger cars due to their stiff suspension and heavier up-sprung mass (for carrying cargo) (cf [Section 6.3.2](#)).

Section 8.3 Performance of the Computational Model

This section describes the performance of the Random Forest computational model in predicting IRI from measured vehicle body accelerations using the dataset prepared above. The performance of the computational model was evaluated by comparing the IRI values predicted by the model for each combination of vehicle type, speed, sprung mass and road profile with the known IRI values.

As described in the Section 2.3.1, the IQL is used as a means of classifying the level of detail required to support various road management activities. Data at IQL-1/2 is regarded as suitable for programming whilst data at IQL-3/4 is suitable for strategic planning.

In order to investigate whether IRI data determined using a smartphone based system could be used for programming and or strategic planning, the Random Forest model was

used to build two separate models as follows:

- a) A regression model to predict IRI at IQL-1/2.
- b) A classification model to predict road condition at IQL-3/4 (i.e. according to the discrete linguistic variables of very good, good, fair, poor, very poor).

In both cases the Random Forest Algorithm was built using the KNIME software (KNIME, 2018). Further details of how Random Forest algorithm was constructed using KNIME are provided in [Appendix A1](#).

Section 8.3.1 Regression Model (IQL-2)

Three performance measures described in the [Section 7.4.1](#) (R-square, Mean Absolute Error and Root Mean Squared Error) were used to determine the accuracy of the prediction of the model. A number of steps were included to evaluate the performance of the computational model as follows:

1. A sensitivity analysis was carried out to help to calibrate the random forest algorithm (i.e. to choose the model parameters).
2. Multiple tests (i.e. in total of 10 tests) were performed using the random forest algorithm, and in each test 500 rows of data were randomly extracted from the dataset (described in Section 8.2.3).

3. Leave-one-out validation was used to evaluate the accuracy of the model when only one dataset is available (i.e. one road user in one vehicle type traveling at one speed on one road sections; See **Table 8.8**).
4. Stratified sampling under different vehicle types was chosen to evaluate the effects of different traffic compositions on the performance of model
5. The results of the performance measures were summarised to give a sense of the overall accuracy of the model together with hypothesis testing to check the statistical significance between the predicted and actual road IRI values.

Random Forest Parameters:

An analysis was carried out in order to set within the Random Forest the most appropriate values of two important Random Forest parameters, the number of trees and the tree split depths (Liaw and Wiener, 2002). To this end, a randomly selected set of 500 vectors were obtained from the dataset established above. The remainder of the dataset (806,844 vectors) was used to build the model. **Table 8.6** presents the results of the analysis.

Table 8.6: Sensitivity analysis

The number of Trees (N)	Tree depths (the number of levels)	Accuracy (R^2)	Time for running
50	Unlimited	0.972	4 minutes
100	Unlimited	0.974	9 minutes
250	Unlimited	0.971	20 minutes
50	10	0.777	2 minutes
50	20	0.953	3 minutes
50	50	0.972	4 minutes
100	10	0.776	2 minutes

From **Table 8.6** above, limiting the tree depth appears to hinder the performance of the Random Forest. Besides, with unlimited tree depths, when the number of tree is above 50, increasing the number of trees in the ‘forest’ does not seems to improve the accuracy of prediction, although it requires increased computational time.

Although there is no clear guidance in the literature on how many trees should be included, it was felt from a scrutiny of the results presented in **Table 8.6** that 100 trees in the ‘forest’ with unlimited tree depth would be an appropriate combination. This combination was used henceforth.

Section 8.3.1.1 Test Results

Table 8.7: Summary of the performance of Random Forest Regression model

Number of tests	Accuracy (R^2)	Mean absolute error (m/km)	Root mean squared error (m/km)	Wilcoxon signed rank test	
				Z-score	P-value (two tailed)
1	0.972	0.229	0.306	2.863	0.004
2	0.975	0.219	0.29	0.345	0.730
3	0.971	0.223	0.309	2.76	0.006
4	0.974	0.218	0.299	1.519	0.129
5	0.974	0.215	0.293	1.461	0.144
6	0.972	0.229	0.304	1.486	0.137
7	0.977	0.213	0.289	1.199	0.231
8	0.969	0.235	0.329	2.219	0.027
9	0.975	0.223	0.301	1.227	0.22
10	0.971	0.24	0.316	0.969	0.333
Average	0.973	0.224	0.303	1.605	0.196

By inspection of **Table 8.7** it can be seen that the model can predict IRI to an accuracy of one decimal place (average Root-Mean-Square error of 0.30 m/km). The average mean

absolute error between the predicted and actual IRI was 0.22 m/km. It can be seen that root mean squared error (MASE) found was generally higher than the mean absolute error (MAE) since the MASE takes into account the weighted score of errors, the larger the error the heavier the weighting.

Wilcoxon signed rank test

To determine the statistical difference, between the predicted and actual road IRI values, the non-parametric Wilcoxon signed rank test was chosen instead of the more usual t-test, because it does not require the data to be normally distributed nor is the data set required to large. The Wilcoxon test assumes the following (Woolson, 2008).

1. The data are paired
2. Each pair is randomly selected and independent

Accordingly, a paired two-sided test was carried out to compare the population means of predicted road roughness and actual road roughness with a desired confidence level of 95%. The hypothesis alternatives were defined as follows:

Null hypothesis= H_0 = differences of population median between the predicted and true IRI equals to 0

*Alternative hypothesis = H_1 = difference of population median between the predicted and true IRI is **not** equal to 0*

The null hypothesis would be accepted if the p -value of the Wilcoxon signed rank test was greater than 0.05 (e.g. 1-0.95) otherwise it would be rejected and the alternative hypothesis would be accepted (<0.05).

The results of the tests are shown in **Table 8.7**, from which it may be seen that although p -value of individual test varies, the average p -value (two-tailed) over these 10 tests is greater than 0.05, and that therefore the null hypothesis can be accepted. i.e. there is no significant difference between the predicted and actual IRI values.

The histogram shown in **Figure 8.4** presents the distribution of the errors between the predicted and actual IRI values for each of the 500 vectors used for the 10 tests. The results show that approximately 68 % of the variance was less than 0.3 m/km, about 25 % of the variance was between 0.3 and 0.5 m/km and 7% could be regarded as outliers (error > 0.5 m/km). It is noticeable that 1%-2% of the errors (about 5-10 points out of 500) were found between 0.8 to 1.3m/km.

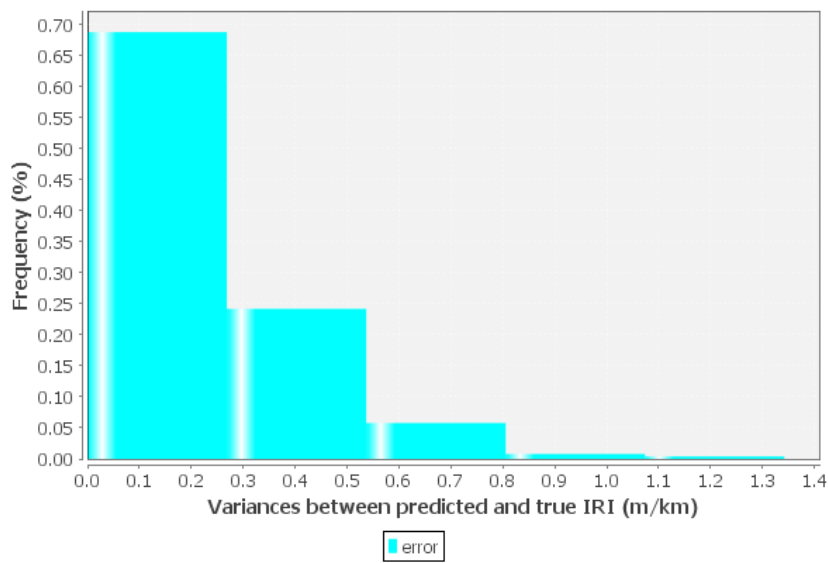


Figure 8.4: *Percentage of the prediction errors between predicted and true IRI using random forest algorithm*

Leave-one-out validation:

A leave-one-out validation process was used to determine the accuracy of the model under the circumstance where a new prediction of IRI is made for any road user using any vehicle type at any speed on any road within the road network. The leave-one-out validation involves drawing one sample of data randomly from the entire dataset in which, except for the leave-one-out sample, all the remaining data are used for training. In this case, the accuracy of model was determined by the one prediction made for the data samples drawn from the datasets. The results of the analysis are shown in **Table 8.8**.

Table 8.8: Leave-one-out validation

Row	N of People	Speed	Vehicle	Grms	True IRI	Predicted IRI	Mean Absolute Error (True vs predicted IRI)
1	1	40	UT	0.195	4.67	5.13	0.399
2	4	50	LEV	0.177	8.07	7.57	0.370
3	1	40	CHB	0.077	7.18	6.92	0.260
4	2	20	BHB	0.026	1.04	1.10	0.194
5	1	30	DMV	0.044	5.24	5.10	0.129
6	2	70	EV	0.18	5.72	5.45	0.099
7	2	40	AHB	0.085	5.49	5.37	0.095
8	1	30	DSU	0.08	5.64	5.22	0.082
9	3	40	LEV	0.064	3.14	3.18	0.041
10	2	20	ESU	0.035	3.38	3.23	0.012
Average:							0.191

As can be seen from the **Table 8.8** the Mean Absolute Error between the predicted and the actual IRI for the 10 leave-one-out validation samples was 0.19 m/km and is consistent with the overall model performance (MAE= 0.22 m/km) discussed above.

The above notwithstanding the variance between the predicted and true IRI was between 0.012 m/km (i.e. ESU) and 0.399 m/km (i.e. UT). Further, from it **Table 8.8** may be seen that the accuracy of the prediction is a function of the vehicle type. Commercial vehicles which include vans and trucks (UT and EV) appear to give a higher error (cf. [Section 6.3.2](#)). In order to investigate this aspect further stratified sampling that enables different proportions of vehicles to be considered within the testing data set was carried out as described below.

Stratified sampling

A sample of 1,000 vectors was obtained from the original dataset through stratified sampling (G. Hall, 2015) by vehicle type. The stratified sampling procedure was such that the number of vehicles within a vehicle type in the sample of 1,000 was the same for all vehicle types. By so doing it enables a better understanding of whether the prediction errors vary by vehicle type. The percentage of all 15 vehicle types of the 1,000 vectors is shown in **Table 8.9** and **Figure 8.5**.

Table 8.9: Summary of stratified sampling in the vehicle types

Category	Percentage of vehicles	Percentage of sample (%)
Hatchback	A (6.9%), B (6.9%) and C-class Hatchback (6.9%) and B-class Sport (3.5%)	24.2%
Sedan	D (6.9%), E (7.0%) and F-class Sedan (6.9%)	20.8%
SUV	D (6.9%), E (6.8%) and F-class SUV (6.8%)	20.5%
Commercial Vehicles	D-class Minivan (6.9%), European van (6.9%) and Large European Van (6.9%), Utility Truck(6.9%) and Pickup(6.9%)	34.5%

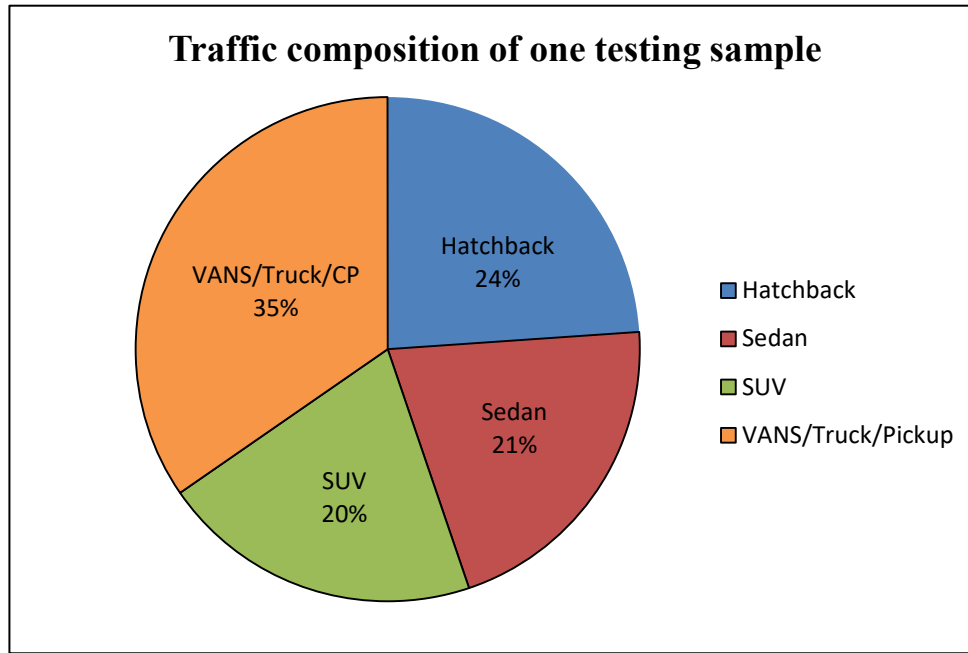


Figure 8.5: The distribution of four traffic categories in the testing sample

The accuracy of prediction, MAE and RMSE for each category of vehicle were then obtained via carrying out one test using the Random Forest algorithm and the results are shown in **Table 8.10**.

Table 8.10: Results of random forest for each category

Category	Accuracy (R^2)	MAE (m/km)	RMSE (m/km)
Hatchback	0.975	0.215	0.292
Sedan	0.973	0.244	0.309
SUV	0.974	0.212	0.295
Commercial vehicles	0.958	0.27	0.364
Full Model	0.97	0.24	0.323

As can be seen from **Table 8.10**, the prediction error associated with commercial vehicles was greater than for passenger cars.

The MAE and RMSE were calculated separately for each commercial vehicle type and the

results are shown in **Table 8.11**. It can be seen from **Table 8.11** that the RMSE of these vehicle types were all higher than the 0.3 m/km average for all vehicle types, with Utility Trucks showing the highest RMSE of 0.43 m/km.

Table 8.11: Prediction errors of Vans/Trucks/CP those commercial vehicle types in the computational model

Vehicle Types	Mean absolute error (m/km)	Root mean squared error (m/km)
Utility Truck	0.336	0.432
Large European van	0.288	0.387
Compact Pickup	0.27	0.37
European Van	0.239	0.313
D-class Minivan	0.218	0.303
Average	0.27	0.364

The above suggests that the commercial vehicles tend to increase the prediction error of computational model as a whole. As demonstrated in [Section 6.3.2](#), comparing to other vehicle types, these vehicles have a relative stiffer suspension and a rigid rear axle, together with a heavier unsprung mass (See **Table 6.7**), large variances can be present in the measured vehicle body acceleration data, particularly for rough roads (i.e. IRI>5 m/km) and when the speed is greater than 50 km/h.

A number of additional tests were undertaken to better understand how the proportion of commercial vehicles in traffic composition may affect the overall model performance. This was achieved by varying the proportion of commercial vehicles in the data sample (i.e. in the randomly selected 1,000 samples). Since originally the proportion of commercial

vehicle in the dataset was 35% (See **Table 8.9**), for comparisons, the proportions chosen in these tests were 0%, 5%, 15% and 50% respectively. The results are shown in **Figure 8.6**.

As can be seen from the **Figure 8.6**, performance of the computational model decreases with the increase in the percentage of the commercial vehicles. For example, the RMSE increased from 0.28 m/km to 0.34 m/km (i.e. by 24%) when the proportion of commercial vehicles was increased from 0% to 50%. The computational model achieves the best performance ($R^2 = 0.979$) when commercial vehicles were excluded (0%).

In addition to above, to show that the decreased model accuracy has a function of increased proportion of commercial vehicles, rather than due to decrease in the proportion of other vehicle types, extra tests were undertaken by eliminating each vehicle group (4 groups in total) in turn then comparing the performance of the models. The results are shown in **Figure 8.7**. By inspecting **Figure 8.7**, it can be seen that by introducing the commercial vehicle group, the performance of the model decreased. Changing the proportion of other vehicle groups within the dataset did not have much impact on the performance of the model compared to changing the proportion of commercial vehicles. This corresponds to the finding summarised in **Figure 8.6**.

As described above, the reason that more commercial vehicles in traffic composition leading to a decreased model accuracy is because there are a combination of contributing

factors causing the large variances of simulated Grms of these commercial vehicles comparing to other vehicle types, which therefore increases the difficulties in the prediction, typically at various of operating speed. These factors included, for example, those vehicle related factors stiff suspension, heavy unsprung mass and various loadings and etc. To illustrate this, **Figure 8.8** compares the distribution of simulated Grms values of 1,000 samples for Hatchbacks and Commercial vehicles. The Grms data for comparison were randomly extracted from the training dataset (0.8 million) with IRI values of selected data were between 3.5 to 4.5 m/km. The statistical summary of these two sets of data (i.e. 1000 each) are shown the **Table 8.12**. As can be seen from these results, a larger variance in the simulated Grms was found for Commercial vehicles compared to Hatchbacks, as evidenced by larger ranges and Standard Deviation.

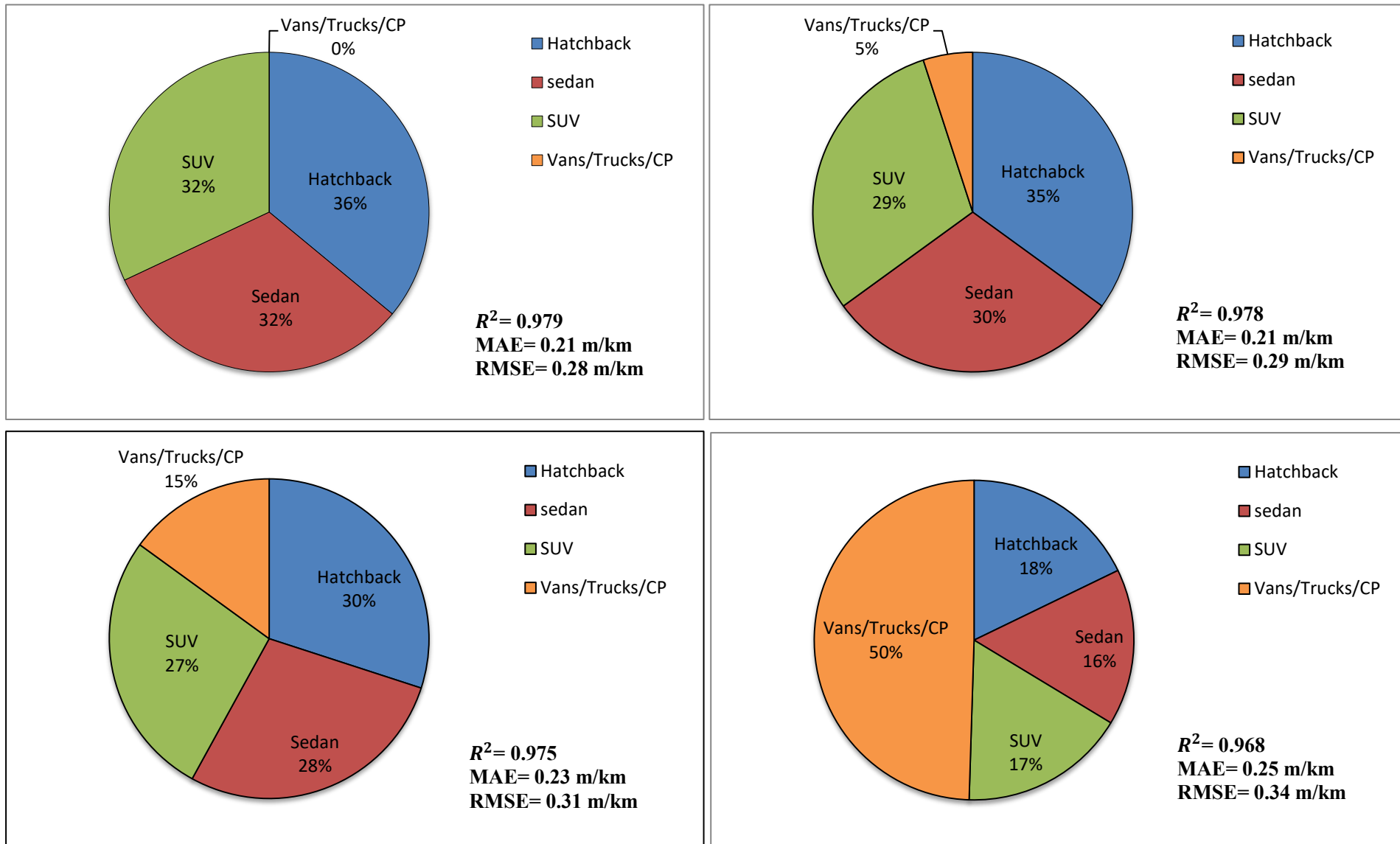


Figure 8.6: The effects of traffic composition (i.e. commercial vehicles) on model prediction accuracy

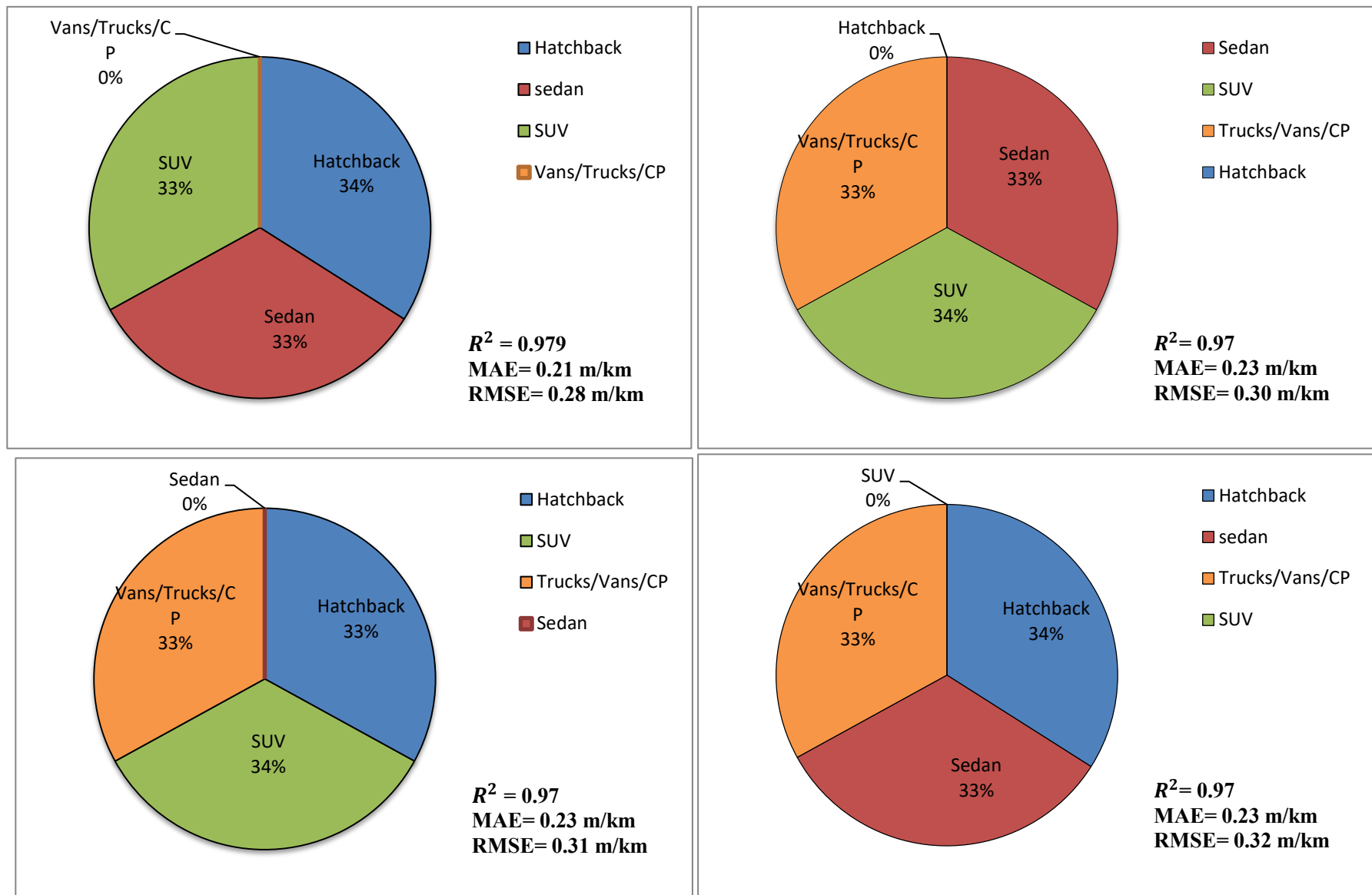


Figure 8.7: The model accuracy of eliminating the vehicle group in turn (4 groups in total)

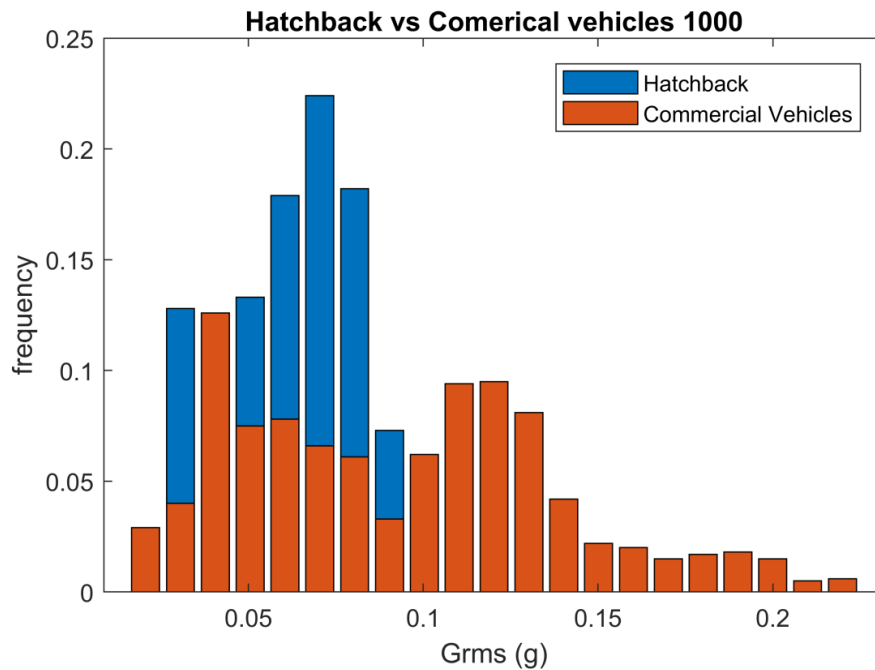


Figure 8.8: Distribution fittings of simulated Grms of Hatchbacks and Commercial Vehicles

Table 8.12: Statistics summary of the simulated Grms (distribution) for Hatchbacks and Commercial vehicles

Vehicle groups	Sample size (N)	Min (g)	Max (g)	Mean (g)	Range (g)	Std.Dev (g)	Skewness	Kurtosis
Hatchback	1000	0.028	0.110	0.065	0.081	0.019	-0.146	-0.561
Commercial vehicles	1000	0.021	0.220	0.093	0.200	0.046	0.464	-0.463

A non-parametric Mann-Whitney U hypothesis test (McKnight and Najab, 2010) was performed to assess whether two independent groups are significant different from each other and results showed that null hypothesis has been rejected ($\text{sig} < 0.05$), which means,

the distribution of Hatchback and Commercial vehicle are different (i.e. two samples are drawn from different population). This also proves that the difference of distribution of two groups (Hatchback and Commercial vehicles) is not due to chance.

Mann-Whitney Test

Ranks				
Group		N	Mean Rank	Sum of Ranks
Hatchback	1	1000	825.64	825640.00
	2	1000	1175.36	1175360.00
	Total	2000		

Test Statistics ^a	
	Hatchback
Mann-Whitney U	325140.000
Wilcoxon W	825640.000
Z	-13.541
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Group

Figure 8.9: Results of Mann-Whitney U test (Group 1: hatchback; Group 2: Commercial vehicles)

From the above analyses it can be concluded that the proportion of commercial vehicle types (UT, CP, LEV, EV and DMV) can affect the performance of the computational model. Furthermore, this would imply that the greater the proportion of commercial vehicles within a fleet of all vehicle types from which vertical acceleration data is obtained to measure IRI, the greater the likely error in IRI prediction (assuming that vehicle type is not accounted for in the prediction model). Conversely it would appear that excluding

commercial vehicles from such a system would provide satisfactory results in terms of IRI prediction using a Random Forest Regression model.

For a commercial system possible solutions to this issue could be to (i) increase the size of the training data and in particular the proportion of commercial vehicle types, and (ii) develop a system that uses only a single or similar type/class of vehicle (e.g. D-class sedan). Approaches to this end are discussed further in Section 8.3.1.2.

Section 8.3.1.2 Single Vehicle Fleet

The above has shown that the selected Random Forest computational model can predict the IRI with reasonable accuracy (i.e. to IQL-2) from a large dataset. Typically, the accuracy was found to be 97.3 % and the average Root-Mean-Square Error was about ± 0.3 m/km, depending on the composition of the vehicles in the dataset.

In practice commercial systems which measure vehicle body accelerations using a smartphone based approach could be employed usefully on a fleet of vehicles of similar vehicle type which transverse large parts of a road network (e.g. taxis, dust carts, police vehicles, ambulances).

The focus of this section is on the accuracy of the chosen Forest Tree in predicting the IRI for vehicle fleets of similar vehicle types. More specifically, does the system perform accurately when it installed in very similar vehicle types to measure the roughness of one particular road section? If yes, how close are the predictions? Also, can a representative

IRI value be determined?

To answer these questions, a dataset was obtained using CarSim™ for which Grms values were obtained from one vehicle type (D-Class Sedan), traversing one road section (IRI=3.12 m/km), travelling at speeds of 50km/h. This combination of vehicle class, road condition and speed was chosen to reflect a taxi driving in urban environment, where the speed limits is 50 km/h and the road condition is good. (See **Table 8.3**).

Accordingly 116 samples were obtained, the dataset was analysed using the Random Forest computational model described above. The results produced by the Random Forest model are shown in **Figure 8.10** in the form of a histogram of predicted IRI values. The statistics for all predicted IRI values are summarised in **Table 8.13**.

Figure 8.10: The histogram plot of testing results (single D-class Sedan fleet)

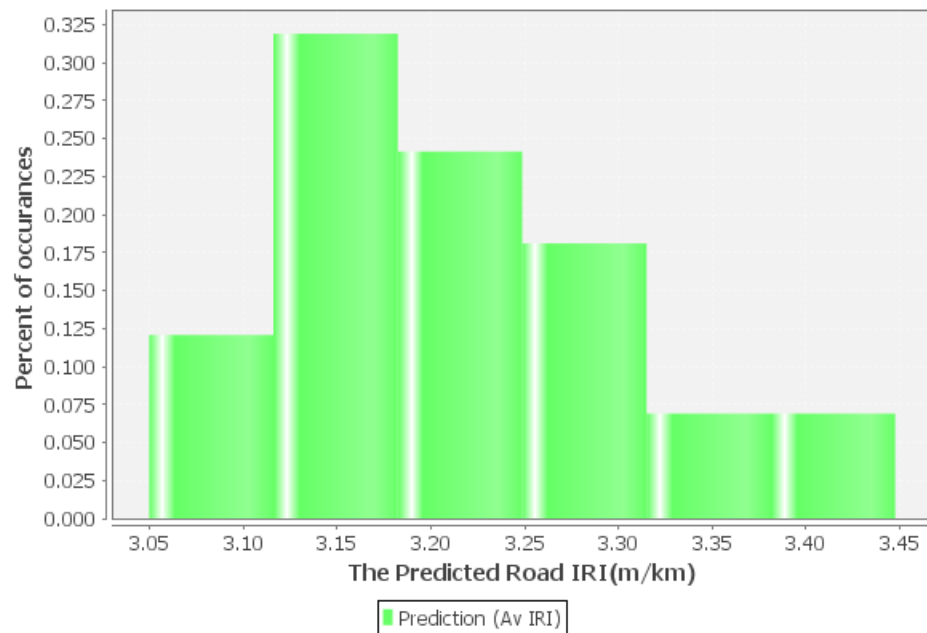


Table 8.13: Statistics of all predicted IRI of testing data (single D-class Sedan)

Mean Absolute Error (m/km)	Root mean Square Error (m/km)	Min	Max	Mean	Median	Std. Deviation	Skewness	Kurtosis
0.097	0.128	3.05	3.45	3.21	3.19	0.091	0.663	-0.234

From **Figure 8.10** and **Table 8.13** it can be seen that the predicted IRI values of all 116 samples lie between 3.05 and 3.45m/km (against an actual IRI value of 3.12 m/km), with mean, median and RMSE values of 3.21, 3.19 and 0.13 m/km respectively. It can be seen that RMSE of the model considering the D-class sedan only is smaller than the overall model involving a number of vehicle types (i.e. 0.30 m/km), suggesting that a commercial system built on these particular vehicle type (e.g. taxis, police cars) may perform better than a system that uses all vehicle types.

From **Table 8.13** it would appear that for this particular case (i.e. predicting IRI from data captured from a number of similar vehicles traveling on the same road section at a similar speed) that the median value would be a better predictor of the IRI than the mean value. This is because the sample size is small and median value takes into account the outliers. However with a large sample size the mean value should be a more reasonable predictor.

Section 8.3.1.3 *Effect of smartphone type*

Thus far, the smartphone-based prototype system is based on the simulated datasets obtained from CarSimTM. However, if real data from the actual smartphones have been used instead of simulated data, the performance of the existing prototype system could be affected (i.e. due to the effect of smartphone type). As mentioned in the

literature review in Section 6.2 and from the practical application presented in **Chapter 5**, smartphones can be sources of several types of error. However, the effect of smartphone associated variability cannot be simulated in CarSim. To address this issue an investigation was carried out to determine, via a sensitivity analysis, the impact of possible smartphone sources of error on the measurement of IRI.

Two types of smartphone associated errors were introduced. These are an equal distribution error and normal distribution of error. These error distributions have been chosen to simulate the practical data collection process whereby different smartphone types are used in which increased variance can be found in the measured vehicle body accelerations. For the former variance type a random error of equal distribution (within $\pm 30\%$) has been added into the CarSim developed simulated Grms dataset. For the latter the error has been assumed to be normally distributed. Datasets in both scenarios were scrutinised as described in **Chapter 8**. The results are given in the following sections.

Errors of equal distribution in simulated Grms

The proposed error of equal distribution was generated using @Risk (see **Figure 8.11**) and the distributions of original Grms datasets (cf. Section 8.2) and new dataset (i.e. Modified Grms_Variance) with added variance (i.e. after introducing a new distribution of error) are shown in **Figure 8.12** and **Figure 8.13**, respectively. **Table 8.14** also summarizes the statistical comparison of the two datasets.

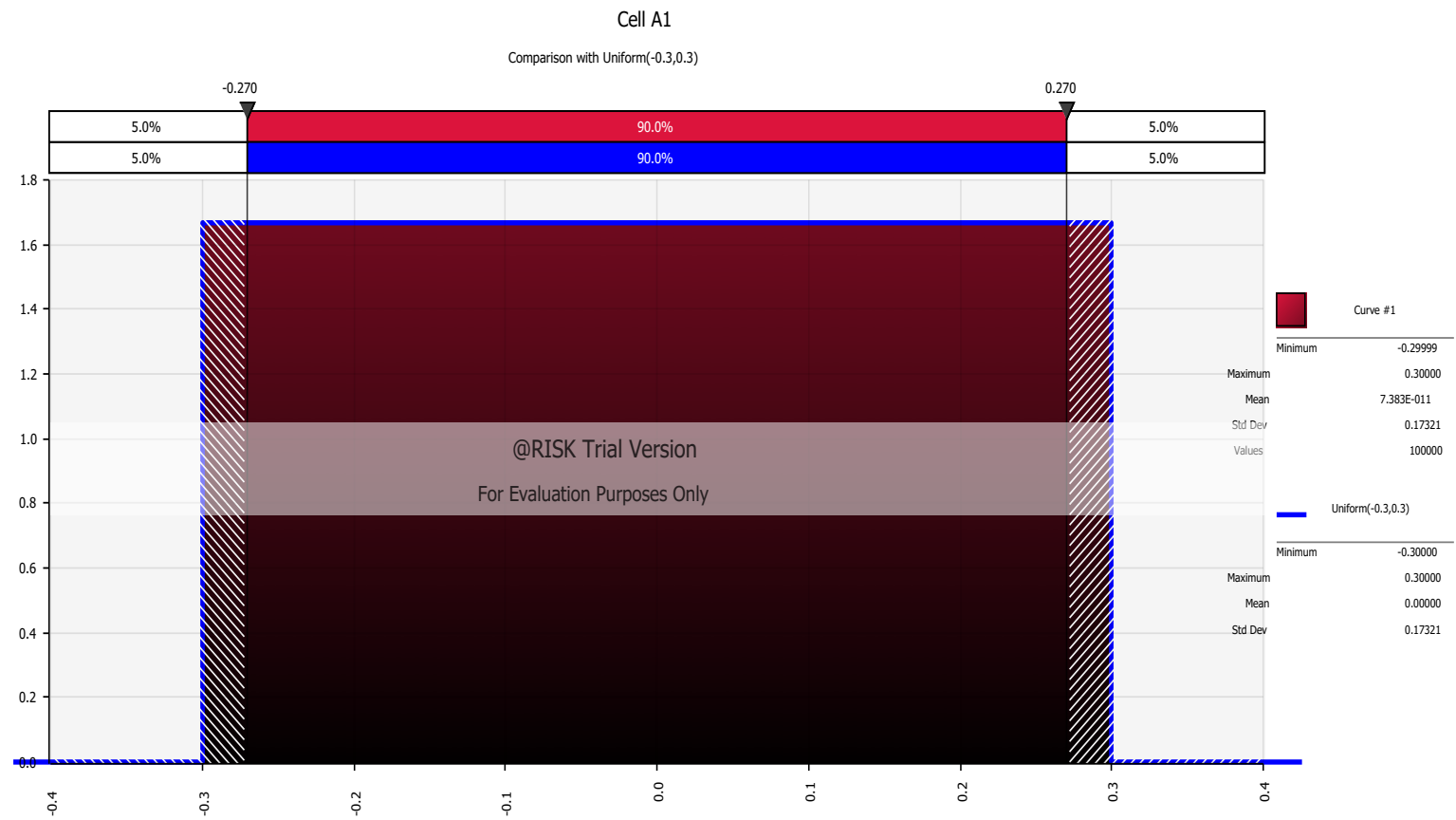


Figure 8.11: Random error with equal distribution modelled using @Risk

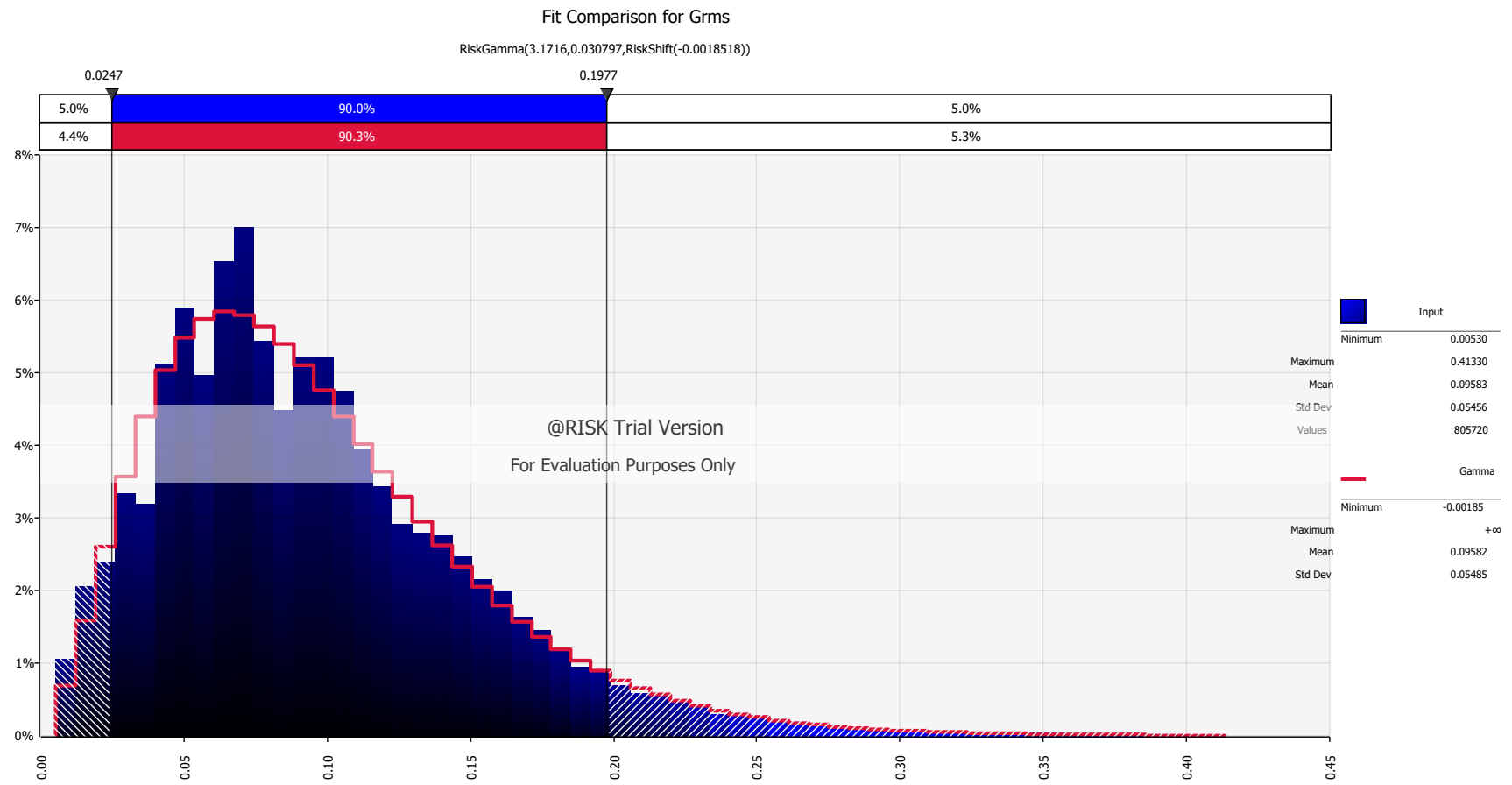


Figure 8.12: Frequency plot of the entire original Grms datasets in @Risk

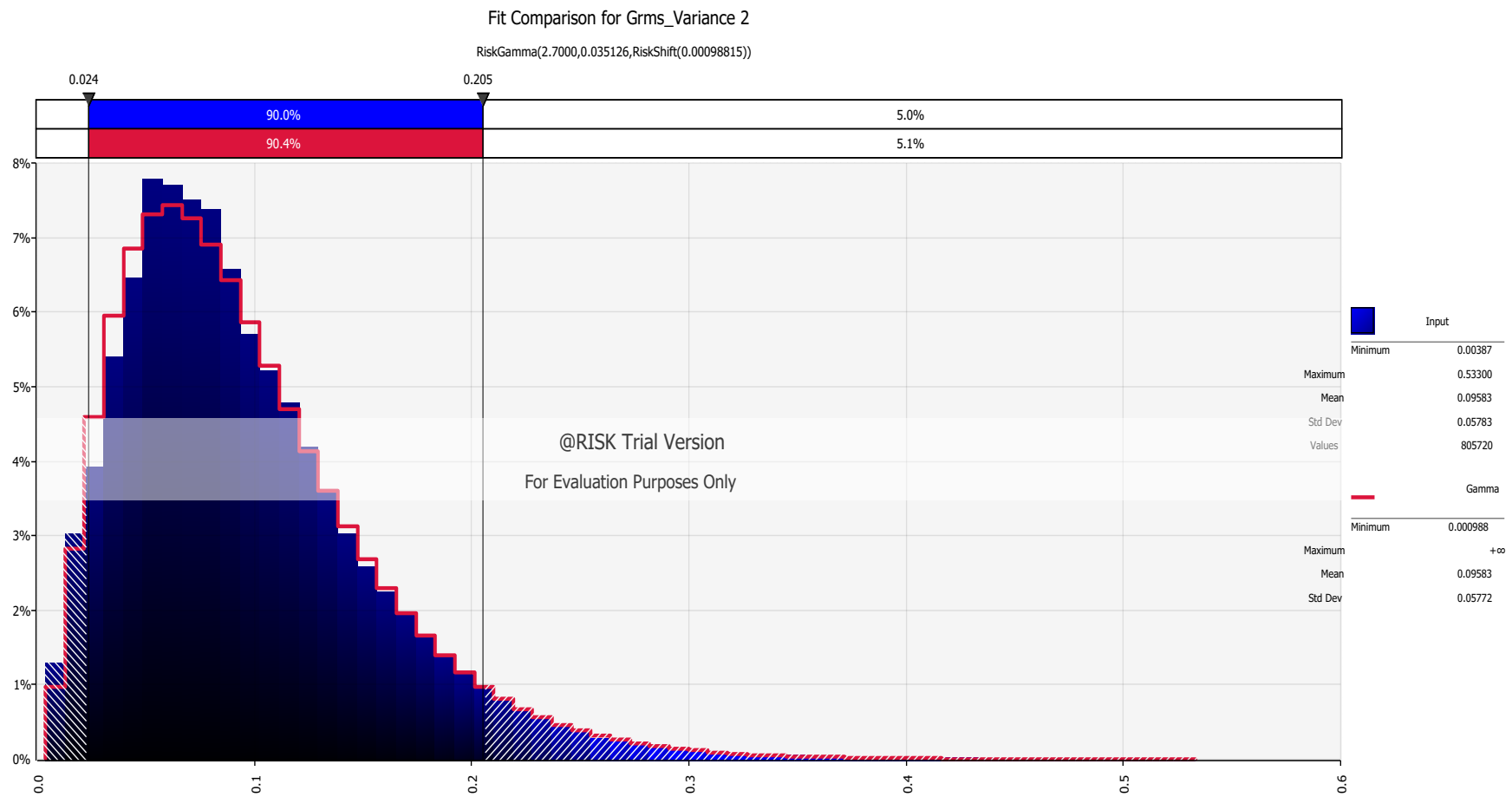


Figure 8.13: Frequency plot of the entire modified Grms_variance datasets in @Risk

Table 8.14: Compare statistics of original Grms and modified Grms_Variance (with errors) Datasets

Column	Min	Max	Mean	Std. Dev	Skewness	Kurthosis
Grms	0.005	0.413	0.096	0.055	1.115	1.882
Grms_Variance	0.004	0.533	0.096	0.058	1.29	5.658

The above figures and tables show that by introducing random error, the distribution of datasets has changed (as evidenced by higher skewness and Kurthosis of Grms_Variance in **Table 8.14**).

After re-testing the prototype system using the modified Grms_Variance dataset for training, the performance of the system is shown in **Table 8.15**.

Table 8.15: Summary of performance of random forest regression using modified dataset

Number of Test	Accuracy (R^2)	Mean Absolute Error (m/km)	Root Mean Squared Error (m/km)
1	0.765	0.72	0.915
2	0.736	0.725	0.939
3	0.757	0.719	0.938
4	0.768	0.717	0.913
5	0.776	0.712	0.914
6	0.782	0.706	0.911
7	0.773	0.712	0.914
8	0.759	0.738	0.944
9	0.775	0.71	0.904
10	0.754	0.721	0.920
Average	0.765	0.718	0.921

Table 8.15 shows that the accuracy of the prototype system has dropped to an average $R^2 = 0.765$ comparing to the $R^2 = 0.973$ (see **Table 8.7**) using original Grms dataset,

the RMSE has also increased to 0.92 m/km compares to RMSE= 0.3 m/km (see [Section 8.3.1.1](#)). This result suggests that the accuracy of the prototype system degraded from IQL-2 (Class 2) to IQL-3 (Class 3) according to **Table 2.4**, which could only be suitable for network level management. This result also show that the effect of smartphone type on the accuracy of the prototype system was significant and should be considered with care during data collection.

Errors of normal distribution in simulated Grms

Because the error of equal distribution assumes different smartphones can add arbitrary error terms (within $\pm 30\%$) to the measured Grms dataset with an equal likelihood, the second error type considered is a normal distribution (see **Figure 8.14**). This consideration is based on the assumption that fewer smartphone types have higher error terms, while most smartphones will perform relatively similarly.

The frequency plot of the modified Grms_Variance (with error) is shown in **Figure 8.15**. **Table 8.16** summarizes the statistical comparison between the original Grms (see **Figure 8.12**) dataset and the modified Grms dataset (Grms_Variance).

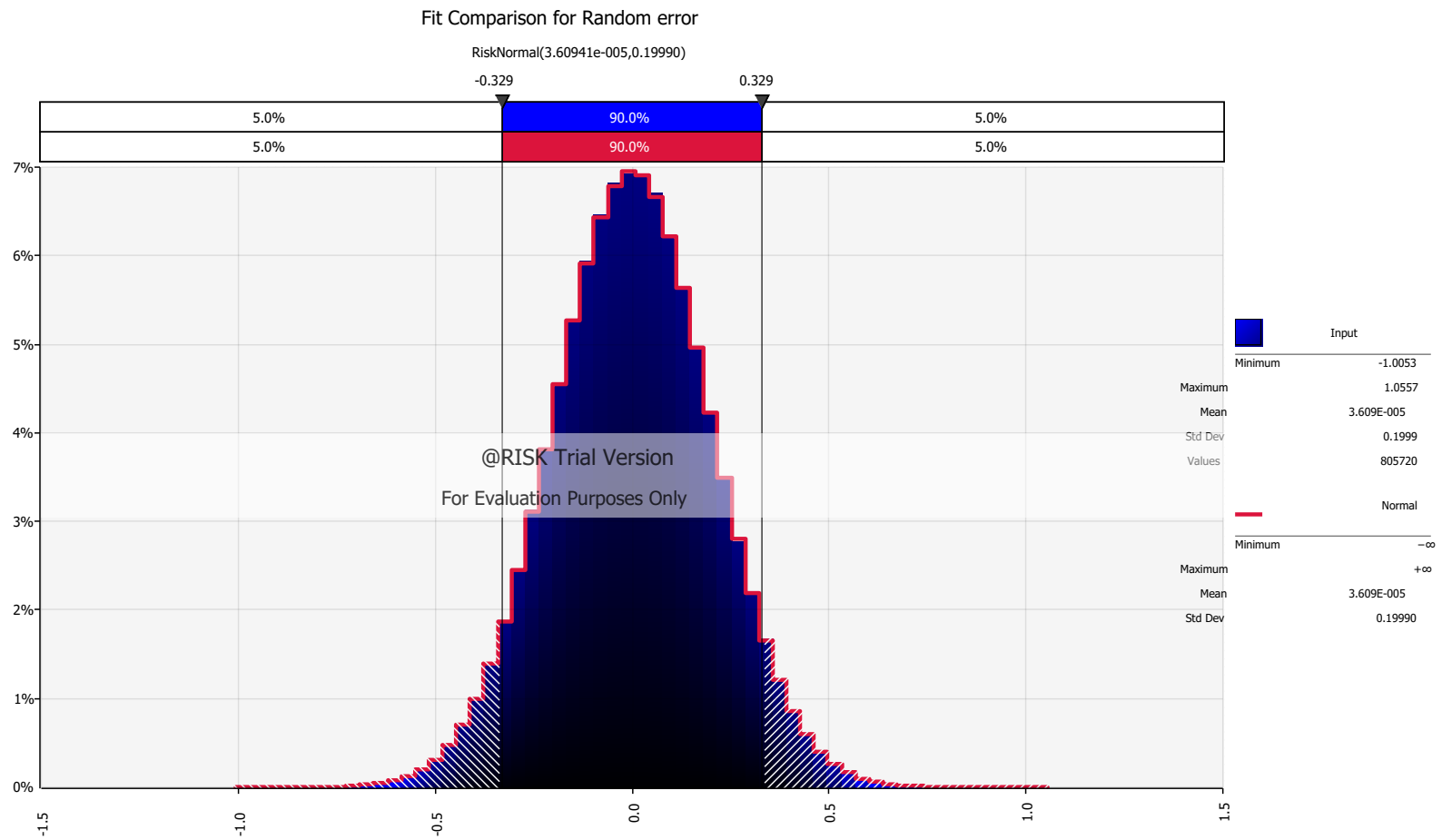


Figure 8.14: Error of the normal distribution modelled using @Risk

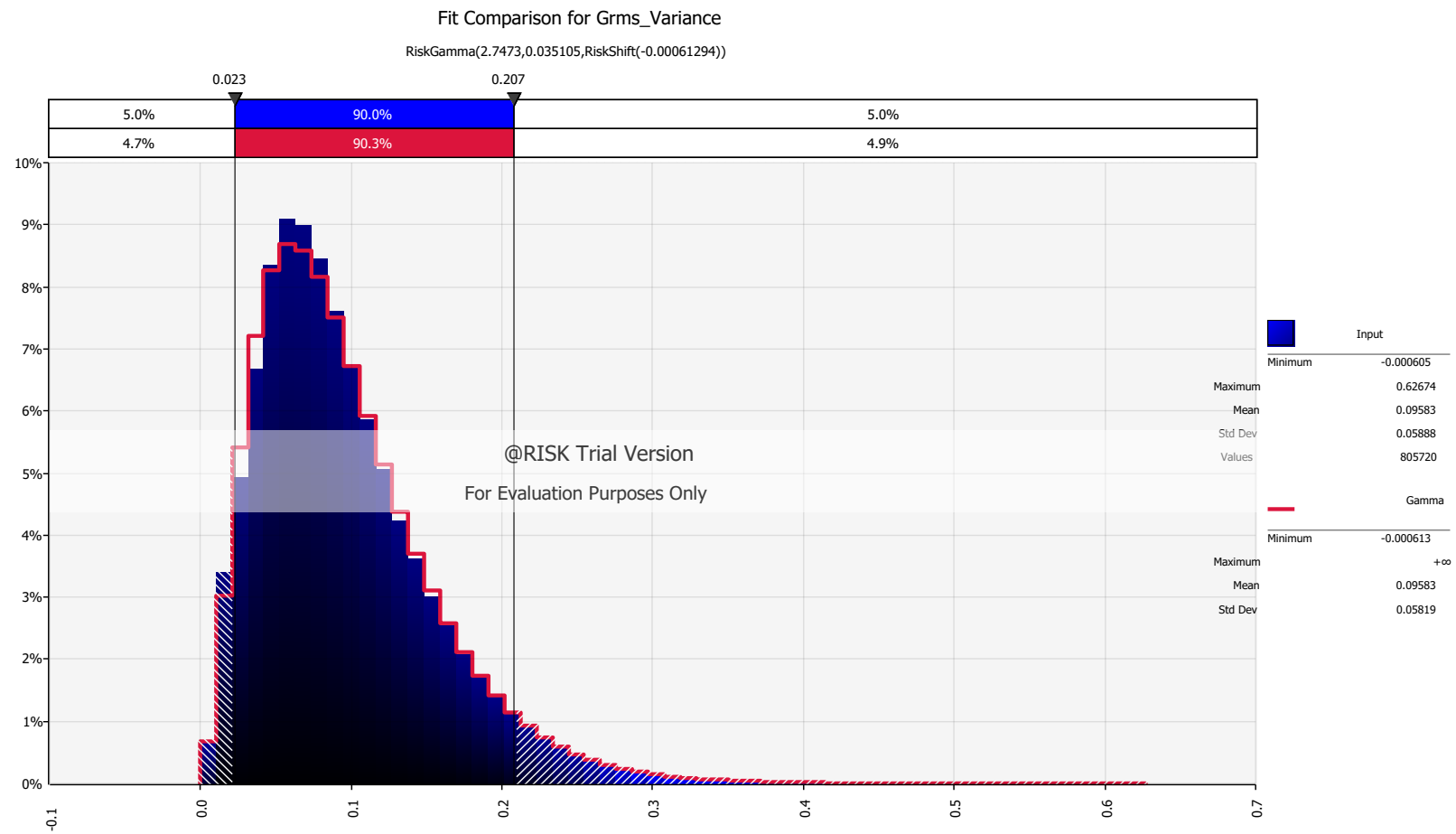


Figure 8.15: Frequency plot of entire modified Grms_variance datasets (with error) in @Risk

Table 8.16: Statistics of the original Grms and modified Grms_Variance (normal distribution) datasets

Column	Min	Max	Mean	Std. Dev	Skewness	Kurthosis
Grms	0.005	0.413	0.096	0.055	1.115	1.882
Grms_Variance	0.0006	0.627	0.096	0.059	1.33	5.93

Table 8.16 shows that the distribution of the new Grms dataset (Grms_Variance) with the introduced error has changed compared to the original Grms dataset, as suggested by the higher maximum value, kurthosis and skewness.

The results following the methodology of Chapter 8 are shown in **Table 8.17**.

Table 8.17: Summary of performance of random forest regression using modified dataset

Number of Test	Accuracy (R^2)	Mean Absolute Error (m/km)	Root Mean Squared Error (m/km)
1	0.68	0.803	1.044
2	0.67	0.803	1.063
3	0.71	0.813	1.036
4	0.68	0.805	1.032
5	0.68	0.818	1.053
6	0.71	0.78	1.02
7	0.69	0.813	1.042
8	0.68	0.825	1.067
9	0.68	0.77	0.985
10	0.67	0.813	1.067
Average	0.69	0.804	1.04

Table 8.17 shows that the average accuracy of the system decreased from $R^2 = 0.973$ (see **Table 8.7**) to $R^2 = 0.69$ and the RMSE increased from 0.3 m/km to 1.0 m/km.

This result causes the accuracy of the system to reduce from IQL-2 (Class 2) to IQL-3 (Class 3) according to the **Table 2.4**. IQL-3 data is suitable for network level analysis.

In order to assess the increased difficulty in predicting IRI once a random error has been introduced to the dataset, the scenario of adding a normal distribution of error data was further investigated. Two sets of 100 data samples were obtained randomly from the entire dataset to represent data collected by two different types of vehicles (Utility Truck and D-class SUV) travelling on roads with different IRIs. A comparison was then visually performed to determine how the Grms values differed before and after introducing the random variable (i.e. effect of smartphone type). The results are shown in **Figure 8.16** and **Figure 8.17**.

The results of Grms_variance (with error) show a different distribution compared to the original Grms according to those two random samples. This difference increased the difficulty for the random forest algorithm to learn the dataset, or making predictions.

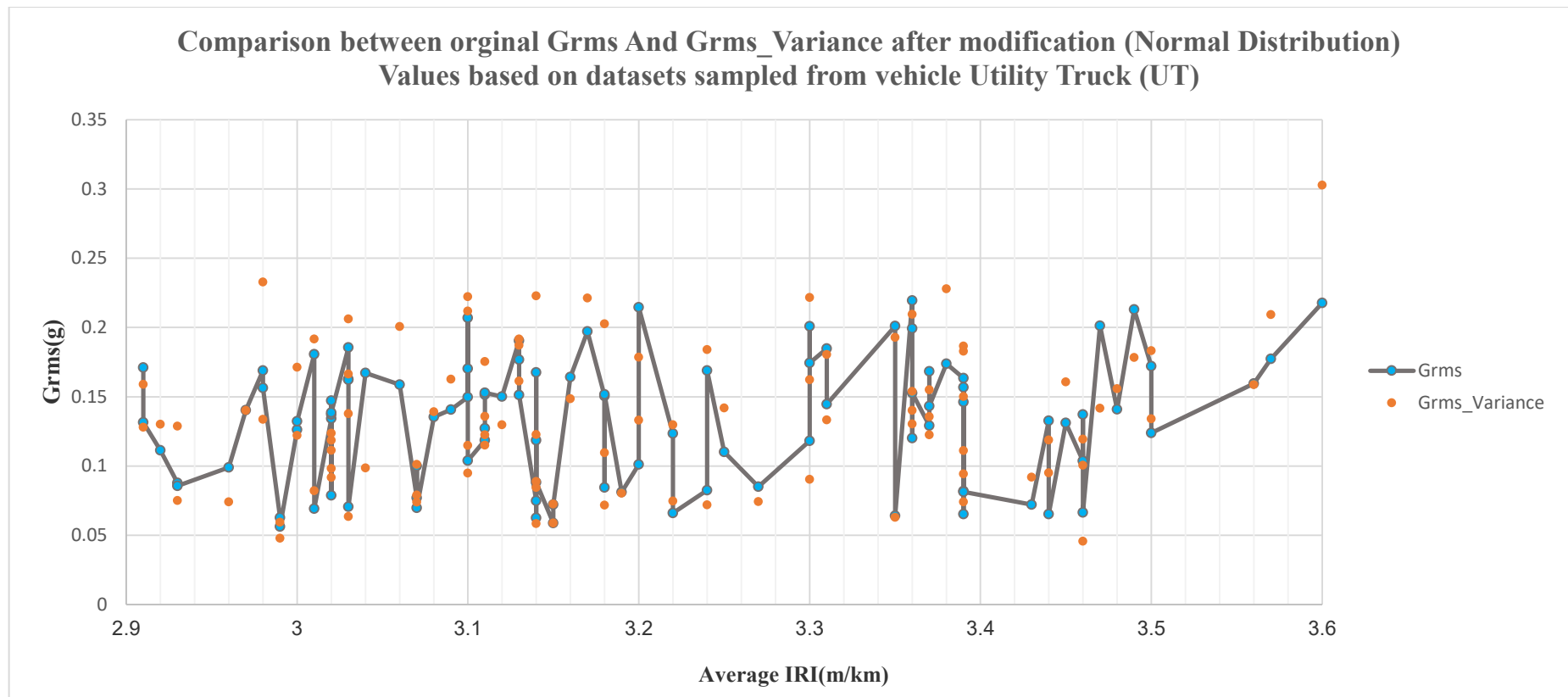


Figure 8.16: Comparison of Grms before and after introducing a random variable (normal distribution)

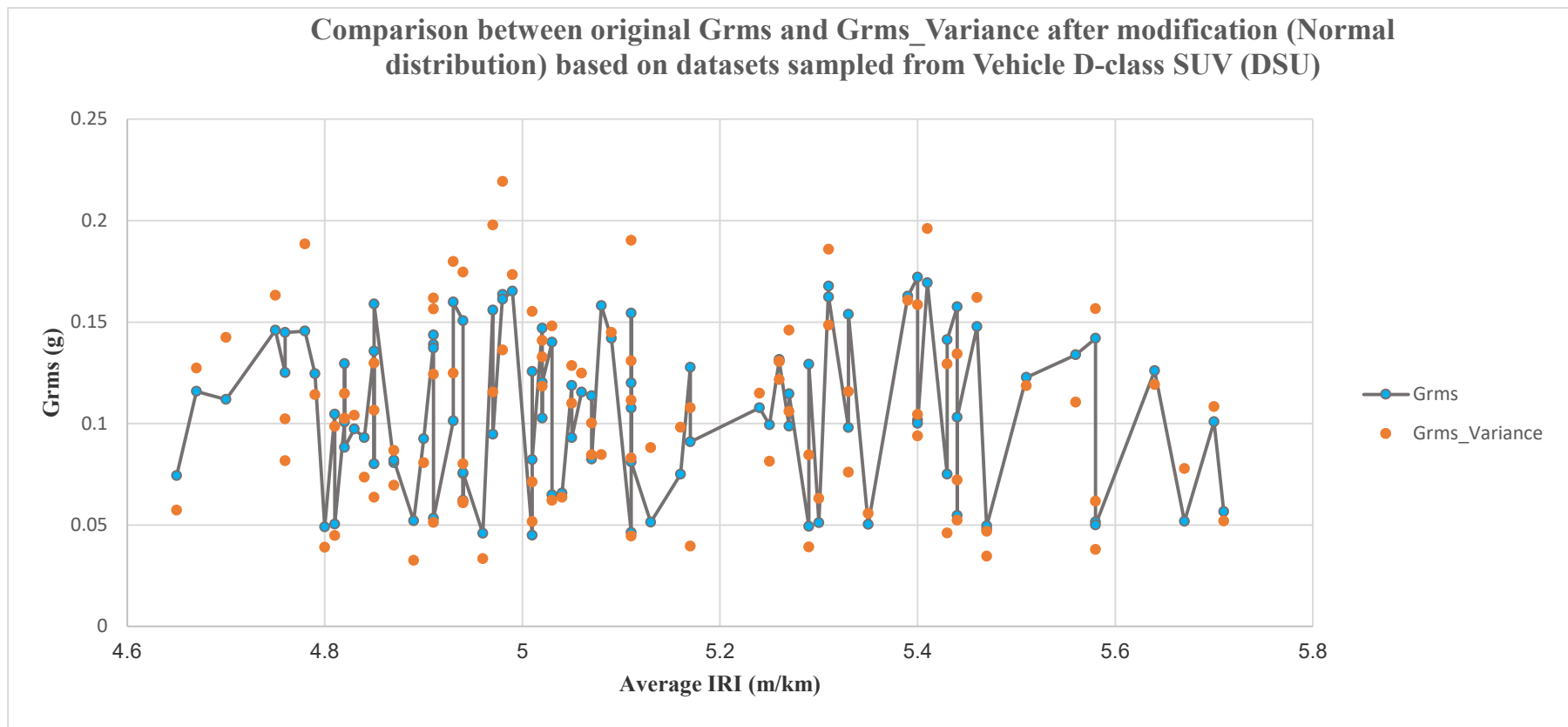


Figure 8.17: Comparison of Grms before and after introducing a random variable (normal distribution)

Section 8.3.2 Classification model (IQL-3/4)

The approach described above was found to be suitable for measuring IRI to IQL-2. However, the amount of data required to ensure reasonable accuracy was very large and required the consideration of a number of features including vehicle speeds and type, the number of people in the vehicle, in addition to the measured vehicle body vertical acceleration RMS values.

In practice, the data associated with these features may not be practicable to capture, or may not be available. Furthermore, as discussed in [Section 2.3.1](#), planning and programming levels of road management decision making (i.e. strategic decision making) only require information at a lower IQL (e.g. IQL 3/4) than that provided by the approach described above. Determining IRI at IQL-3 or 4 can be regarded as a classification problem (i.e. classifying the data into a range of values each of which represents very good, good, fair, poor, and very poor conditions). Accordingly, this section investigates firstly, the performance of a Random Forest algorithm with respect to such a classification problem. Second, the section considers whether it is possible to reduce the number of variables required to be taken into account when determining IRI and at the same time still produce information which can be considered to be to IQL-3 or 4. For the purposes of this research the range of values / classes were defined as shown in **Table 8.3**.

Addressing these classification problems involved:

1. Using an ordinal scale to classify road roughness (IRI) into classes to assess the overall performance of random forest classifier
2. Using a feature elimination process to find out the relative effects of variable(s) in the

dataset on the accuracy of the prediction of IRI.

3. Eliminating the predetermined feature and carrying out a new prediction.

Section 8.3.2.1 Random forest classifier

Four performance measures described in the **Chapter 5** (accuracy, precision and recall, and F-score) were used to determine the accuracy of classification. Road roughness (IRI) was classified into different classes according to **Table 8.3**. A total of 10 tests were performed using the random forest algorithm, and in each test 500 rows of data were randomly extracted from the dataset (as described in [Section 8.2.3](#)) for the purposes of evaluating the overall performance. The results are shown in **Table 8.18**.

Table 8.18: Summary of the performance of Random Forest Classification model

Number of tests	Accuracy (%)	Precision (averaged)	Recall (averaged)	F-score (averaged)
1	98.4%	0.982	0.991	0.986
2	98.8%	0.986	0.992	0.989
3	98.6%	0.986	0.991	0.988
4	99%	0.995	0.987	0.991
5	98.6%	0.991	0.985	0.988
6	98%	0.987	0.979	0.983
7	98.5%	0.982	0.992	0.986
8	98.1%	0.980	0.988	0.984
9	98.6%	0.982	0.992	0.987
10	98.7%	0.987	0.998	0.989
Average	98.5%	0.986	0.990	0.987

By inspecting **Table 8.18**, it can be seen that random forest can perform the classification problem well, as evidenced by high accuracy (average 98.5%) of correct classification and the F-score (0.987). Through introducing a classification system which categorises IRI in terms of a range of IRI values, rather one which attempts to predict the IRI exactly

obviously simplifies the problem, reduces the difficulty in prediction and reduces the information quality (from IQL-2 to IQL-3/4). For such a system it is useful to find out whether the number of variables required to be considered when determining IRI to IQL-3/4 can be reduced. To answer this question, a process of feature elimination was used as described below.

Section 8.3.2.2 Backward Feature Elimination

The backward feature elimination loop was used to determine the relative importance of each input in the prediction of IRI (KNIME, 2015). Backward feature elimination involves, at a given iteration, training the chosen Random Forest regression algorithm using all input variables (n). Then one input variable is removed at a time and the same Random Forest algorithm is trained on $n-1$ variables n times. Thereafter the input variable which produces the smallest increase in the error rate is removed, leaving only $n-1$ variables. The process is then repeated using $n-2$ features $n-1$ times. The procedure continues until only one feature is left and the total number of iterations is calculated as $n*(n+1)/2-1$.

The process of the backward feature elimination carried out in the research using the Random Forest algorithm is shown in **Figure 8.18** and summarised in **Table 8.19**.

Figure 8.18: The process of backward feature elimination loop in the model

Iteration 1:

Number of feature	Features included	Accuracy (R^2)	Root Mean Squared Error (m/km)
4	Vehicle, Speed, Grms and N of People	0.971	0.32



Iteration 2:

Number of Feature	Features included	Accuracy (R^2)	Removed Feature	Root Mean Squared Error (m/km)
3	Speed, Vehicle, Grms	0.8	N of people	0.91
3	Speed, N of People, Grms	0.517	Vehicle	1.27
3	Speed, N of People, Vehicle	0	Grms	1.84
3	N of People, Vehicle Grms	0.494	Speed	1.30

By observing the results, the effect of 'People' was found to be the smallest in terms of the increase in error rate, therefore the feature 'People' was discounted, leaving the 3 features of speed, vehicle and Grms



The first feature 'N of People' is removed due to relative least importance

Iteration 3:

Number of Feature	Features included	Accuracy (R^2)	Removed Feature	Root Mean Squared Error (m/km)
2	Vehicle Grms	0.614	Speed	1.14
2	Speed Grms	0.561	Vehicle	1.21
2	Speed Vehicle	0	Grms	1.84

In the second iteration the effects of 'Speed' was observed to be the smallest in terms of the increase in error rate, and it was therefore discounted from further processing. Leaving the 2 remaining features of vehicle and Grms

The second feature 'Speed' was removed due to its less relative importance

Iteration 4:

Number of Feature	Features included	Accuracy (R^2)	Removed Feature	Root Mean Squared Error (m/km)
1	Grms	0.448	Vehicle	1.36
1	Vehicle	0	Grms	1.84

Consequently, the effects of 'Vehicle' was the observed to be the smallest in terms of the increase in error rate, therefore the feature 'Vehicle' was removed, leaving only the feature 'Grms'.

The relative importance of the inputs into the Random Forest Classifier are summarised in **Table 8.19**, where the four independent variables are ranked from 1 to 4, indicating their relative importance from the greatest to the least. The feature 'N of people' appears to have the least impact on the prediction of IRI among the four variables. As shown in **Figure 8.18**, removing the feature 'N of people' (represented by changes to the vehicle sprung

mass) reduces the accuracy of the model from 97.5% to 80% and the Root Mean Squared Error increases from 0.32 m/km to 0.91 m/km.

Table 8.19: the relative important of independent variables

Ranking	Features	RMSE (m/km)
1	Grms	1.84
2	Vehicle	1.36
3	Speed	1.14
4	N of People	0.91
Note: The ranking from 1 to 4 indicates the relative importance of inputs from the largest to smallest		

For the purposes of demonstrating the new classification model particularly where the variable ‘N of people’ was eliminated, the dataset described in [Section 8.2](#) was used, with the following modifications:

1. The input feature ‘N of people’ (i.e. number of people in the vehicle) was removed as it was found, according to the above analysis, to have the least effect on the accuracy of the prediction of IRI by the Random Forest algorithm.
2. Road roughness (IRI) was classified into different classes according to **Table 8.3**.

The Random Forest classifier was used to predict the class of IRI following the same procedures as for the regression analysis described above. Namely, 500 testing data vectors were randomly selected and used to evaluate the performance of the algorithm. The tests were repeated 10 times and the results averaged to account for sampling. In this case, the performance measures chosen to determine the accuracy of predictions were recall (i.e. what percentage of actual positive tuples is classified correctly), precision (i.e. the percentage of predicted positive tuples is true) and the F-score (See [Section 7.4.1](#)). The results from the analysis are shown in **Table 8.20**.

Table 8.20: The accuracy of random forest classifier

Number of test	Accuracy	Recall	Precision	F-score
1	90.6%	0.949	0.895	0.917
2	86.6%	0.934	0.848	0.879
3	83.6%	0.932	0.825	0.857
4	86.8%	0.857	0.934	0.883
5	85.6%	0.928	0.851	0.876
6	87.8%	0.942	0.864	0.890
7	87.6%	0.938	0.866	0.891
8	89.8%	0.947	0.886	0.909
9	84.0%	0.920	0.829	0.859
10	86.2%	0.934	0.862	0.884
Average	86.9%	0.935	0.859	0.885

The Confusion Matrix (introduced in [Section 7.4.1](#)) of a particular example (i.e. Test 4 in **Table 8.20**) from the above model is shown in **Table 8.21**.

Table 8.21: The Confusion matrix (Test 4)

		<i>Ture Condition</i>			
		Good	Fair	Very Good	Poor
<i>Predicted Condition</i>	Good	80	22	0	0
	Fair	5	238	0	0
	Very Good	0	0	39	0
	Poor	0	39	0	37
<i>Correctly classified:434</i>		<i>Wrongly classified:66</i>			
<i>Accuracy: 86.8%</i>		<i>Error: 13.2%</i>			

By comparing the results achieved by the Random Forest classifier described in this section with that of the Random Forest regression algorithm in the feature elimination process (See **Figure 8.18**), the accuracy of model, measured in terms of R-squared, has decreased from 98.5% when all variables were included to 87% when the variable ‘number of people’ was removed. This finding is similar to that of Schlotjes et al. (2014), who undertook an empirical analysis of the performance of a smartphone-based system. Schlotjes et al. (2014) found that the system they tested was able to achieve an accuracy in

measurement of $\pm 20\%$ of actual IRI to IQL-3/4. In addition to this, when the variable ‘number of people’ is eliminated from the regression and classification models, the accuracy of the regression model is 80% (See **Figure 8.18**) compared to 87 % in classification model. This suggests that the classification model performs better than the regression model when less variables are available. However the information quality level for the classification model is IQL-4 whereas that for the regression model is IQL-2/3.

Section 8.4 Summary

In this chapter the accuracy of a Random Forest computational model for predicting IRI values obtained by a smartphones fitted within a fleet of different vehicle types has been assessed. For predicting roughness to IQL-2 a Random Forest regression model was developed and a Random Forest classification model was built to compute IRI to IQL-3/4.

According to the analysis presented in the chapter, the most appropriate parameters for the Random Forest regression model were found to be 100 trees with an unlimited tree depth. Using an artificial dataset generated from a simulation undertaken in CarsimTM, which included consideration of the vehicle speed, type, Grms, Sprung mass and smartphone type, IRI could be predicted within an error of ± 0.3 m/km. The results presented suggested that the computational model is able to predict IRI to an accuracy of one decimal place. This level of accuracy is suitable for IQL-2 measurement (see **Table 2.4**). A system based on the analysis of smartphone data obtained from moving vehicles which can perform to this level of accuracy may be regarded as a Class 2 device. Such a system was also found to satisfy the specification defined for calibrating a vehicle response type (high reliability) road roughness measurement at network level i.e. a standard error of 0.35 m/km (Archondo-Callao, 2009).

In addition, vehicle types and speed were considered as non-negligible variables in the data collection process. The errors in predicting IRI from a dataset containing multiple vehicle types was related to the proportion of commercial vehicles (i.e. Utility Truck (UT), Compact Pickup (CP), Large European Van (LEV), European Van (EV) and D-class Minivan respectively). Considering a model which aims to predict IRI from a single vehicle fleet traveling on the same road section at similar a speed, it was found that the median value would be a better predictor of IRI than the mean value. For IQL-3/4 data (i.e. that required for network-level planning or programming), the Random Forest classifier was found to perform well.

Chapter 9

DISCUSSION

Section 9.1 Introduction

This research investigated whether it is possible to use smartphones fitted inside moving vehicles to collect vehicle body accelerations caused by road irregularities as a means of assessing road conditions. The research also demonstrated how road condition data simulated from such a system can be used in the HDM-4 decision support tool to set economic road maintenance standards (See **Appendix B**). To this end, the software necessary for such a prototype system was demonstrated. A number of objectives were set in the research in order to achieve the stated aim (see Section 1.2). These objectives are further discussed in in this chapter to determine whether or not they have been fully addressed.

Section 9.2 Research Objectives

Chapter 2 addressed objectives one and two of the research (i.e. data requirement for asset management decision making), via an extensive review of the literature to determine the primary measures of road condition, their associated data requirements, and whether it might be possible for a smartphone to collect these primary measures. To this end, the literature review suggested that it might be possible to use smartphone technology to measure road roughness, structural strength and road friction. As demonstrated by this research, it is possible to measure road roughness (cf. Section 2.3.3) using a smartphone fitted to a moving vehicle via collecting vehicle body acceleration. This is also recommended by observations described in **Chapter 5** although the practical application

carried out was not considered as a controlled experiment. In addition, it may also be feasible to measure road friction and structural strength using such a system. This is discussed further in section 9.3.

Chapter 6 addressed objective three (i.e. understanding various influencing factors) via an extensive literature review and the use of a vehicle dynamics package for carrying out a variety of simulations and data modelling processes in order to determine the relative effects of a variety of factors on the measurement of road roughness. Amongst these factors there were a number the impact of which might not be easy to predict because they can occur at any time. Some of these factors include low vehicle operating speed, vehicle cornering and road-profile associated factors, such as road bumps. These aspects are discussed further in section 8.4. In addition, the research found that smartphone type is an important factor affecting the measurement of road roughness and although it is recommended that further experimental work is carried out to quantify the potential affects (see Section 10.2).

Using simulated datasets, **Chapters 7 and 8** demonstrated how machine learning algorithms can be used to process vehicle body acceleration data captured using a smartphone to infer road roughness thus addressing objective four of the research. **Chapter 7** reported on a multiple criteria approach devised to select the most appropriate machine learning algorithm of those identified in the literature review as being potentially useful for the task in hand. Although the multi-criteria approach adopted (i.e. pairwise comparison) is widely reported in the literature it could still be regarded as being subjective, and the weights so determined rely on user/ expert judgement. It might be possible to refine the weights allocated herein by engaging more experts and this could

impact the chosen algorithm (i.e. the Random Forest).

Nevertheless, through the analysis presented in **Chapter 8**, it was found that by using the selected algorithm it was possible to measure road roughness to IQL-2, with an average error of less than ± 0.3 m/km, given information concerning major variables such as vehicle speed and type, sprung mass and Smartphone type. In practice, the accuracy may not be at the same level because a real database would be more complex and contain larger variances than those evident in the simulated dataset used to train any algorithm (see section 8.3.1.3). For example, a number of model variants in one vehicle type, various smartphones being used and a series of different average vehicle speeds can lead to increased variance in the datasets. To address this, machine learning algorithms were trialled to determine whether it might be possible to categorise road condition into roughness bands (i.e. to IQL-3/4), given only vertical vehicle acceleration data (i.e. Grms), vehicle type and speed. This process showed that such a procedure could accurately so quantify road condition.

Section 9.3 Limitations of Research

The fundamental research in this thesis has attempted to determine whether and how a smartphone can be used as collect data to asses road condition. The main limitation of this research is that the prototype system is built upon simulated data generated from the CarSim software. This resulting data may not exactly represent real data measured using smartphones fitted in a moving vehicles. Although the CarSim has been verified by several studies (Organiscak, 2014; Varunjikar, Vemulapalli and Brennan, 2012), CarSim should be considered a nonlinear empirical model (i.e. using sophisticatedly mathematic formulas) that relatively closely replicates the actual vehicle behaviour.

Because the machine learning algorithm has powerful computational capability with a large training dataset (such as those described in section 8.2), the machine learning algorithm predicts the input IRI accurately (as described in [section 8.3.1.1](#)). When transferring the prototype system into real application, where collected real datasets can have larger variance than the simulated datasets from CarSim (i.e. with different distribution), the performance of prototype system may decrease. This was modelled and discussed in [section 8.3.1.3](#).

It is necessary to first determine to what accuracy a smartphone- based system should achieve. Thus far, the lesson learnt from the development of the prototype system suggests that realistically a smartphone-based system can be used to collect to IQL-3 data. In an ideal situation (e.g. under laboratory conditions) such a system may be capable of measuring IRI to IQL-2. The research suggests that an error of $\pm 1\text{m/km}$ IRI can expected.

Section 9.4 Road Friction and Structural Strength

Road friction measurement

While the research has primarily focused on the development of a prototype system for measuring road roughness, this approach could potentially be used to measure road friction. The term ‘road friction’ is used herein to refer to a simplified measure of skid resistance (e.g. classification of the surface skid resistance into very good, good, fair, poor or very poor).

Research has found statistically significant evidence that crashes increases in wet condition as a result of decreased road friction (J. Hall *et al.*, 2009; Pisano and Goodwin, 2003). Skid resistance — the a measure of road safety for driving a vehicle on the road — plays a

significant role in reducing accidents, especially in wet conditions, and can provide a vehicle with the ability to safely accelerate, corner and brake (Mataei *et al.*, 2016). It is described as a limiting coefficient of friction, μ , between a road surface and the vehicle tyre, which is a ratio between the limiting horizontal force (i.e. resisting the braking, accelerating, cornering forces) to the vertical load of the vehicle applied to the tyres.

A vehicle will skid when, in accelerating, braking or cornering, the frictional force required exceeds the maximum friction force that is generated at the tyre-road interface (C. R. Bennett *et al.*, 2007). The maximum frictional force generated is in a position where the status of a vehicle is between skidding and not skidding. The skid resistance is normally measured by calculating the resistance of a test tyre on a wetted road, where the tyre-road friction coefficient, μ , varies between zero and one, depending on the road surface condition considered, such as ice, snow covered, gravel and/ or dry asphalt.

To explore the measurement of skid resistance using a smartphone application, research has shown that vehicle stopping distance correlates to the skid resistance of the wetted road (Greibe, 2007; Heisler, 2002). A standard testing procedure for measuring vehicle stopping distance to determine skid resistance can be found in ASTM E445, “*Standard Test Method for Stopping Distance on Paved Surfaces Using a Passenger Vehicle Equipped With Full-Scale Tyres*”. Accordingly, knowledge of the vehicle stopping distance is used as an additional tool (i.e. proxy) to characterise road skid resistance. This is because the built-in three-axis accelerometer in the smartphone captures the vehicle’s longitudinal deceleration while braking, together with the total time required for the vehicle to stop from a known vehicle operating speed, which then enables calculation of vehicle stopping distance. In addition, the GPS function built into smartphones can be used

to calibrate measured stopping distance, depending on the accuracy of the GPS (e.g. 3 m).

Examining measured vehicle decelerations can help determine whether or not a driver is performing an emergency stop, as stopping distance can be correlated with skid resistance when they are captured under an emergency stop. Multiple confirmations by other road users (i.e. crowdsourcing) are required at the same location, with data sent to a cloud server for computing in order to avoid erroneous classification.

It is notable that there are a number of factors that can affect the measured stopping distance using this approach, including: a) vehicle-related factors (i.e. vehicle speed, performance of the anti-lock braking system (ABS), vehicle type and weights, tyre pressure and tyre thread); b) road-profile-related factors (i.e. road type, road gradient and road friction); and c) human-environment related factors, including driver behaviour (i.e. pedal force, wind drag and wet or dry conditions) (Greibe, 2007). Accordingly, an approach similar to that suggested herein for assessing road roughness could be adopted, namely, an investigation of the predominant factors influencing road friction measurement allied to suitably selected and trained machine learning algorithms to process large datasets for identifying road sections with low friction. **Figure 9.1** illustrates the proposed approach.

Note that a skid resistance measuring system based on the above, could be particularly useful at traffic junctions, approach lanes to a roundabouts and pedestrian crossings, where drivers are most likely to perform an emergency stop. Such information might be particularly useful in a local road context where skid resistance is less likely to be measured on a frequent basis compared to high volume roads.

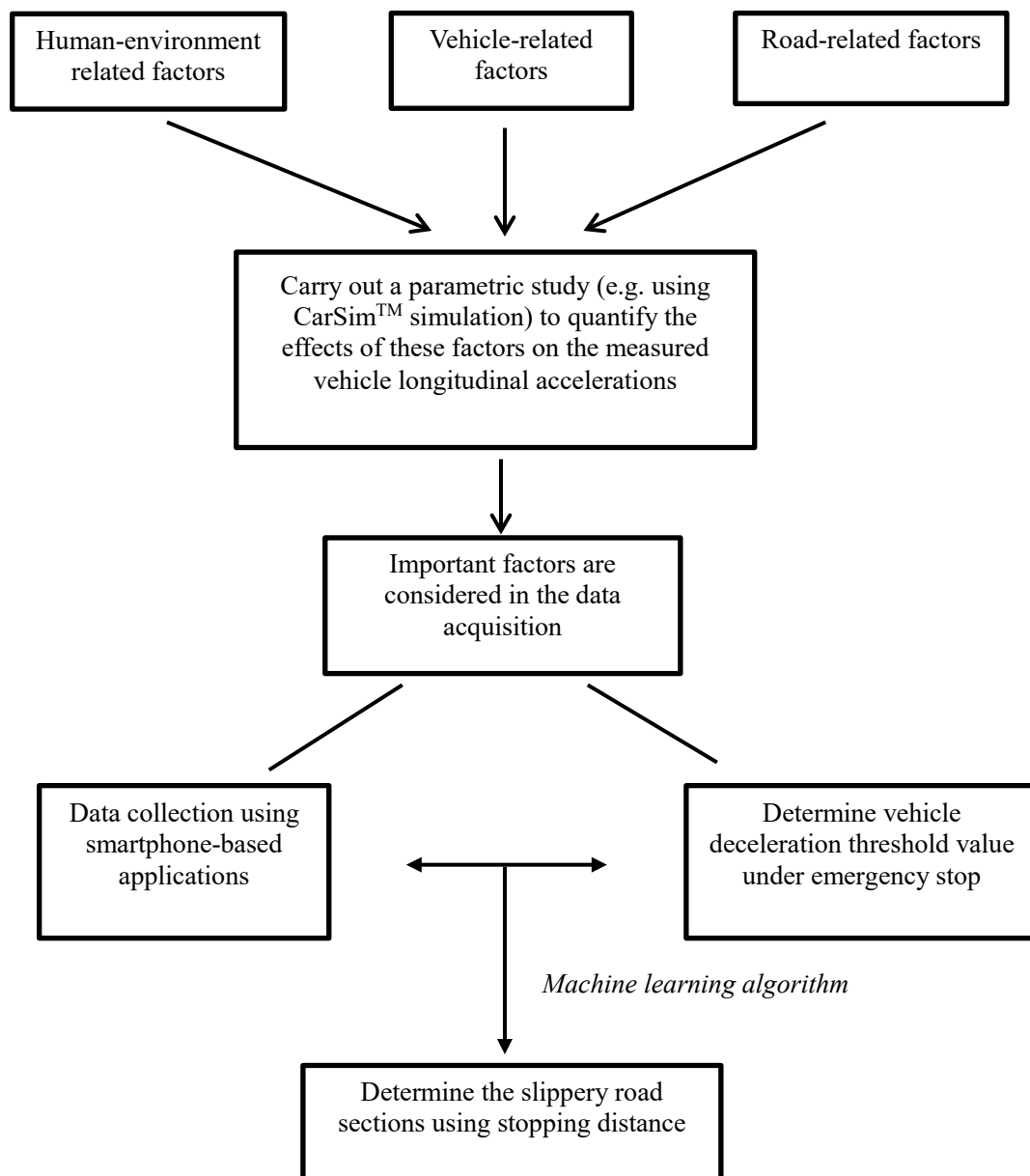


Figure 9.1: Proposed approach for measuring friction on road

From the perspective of the road agency, detecting roads with insufficient friction could be used to determine maintenance and improvement needs. Such information conveyed to the road user might also improve road safety by warning the road user in advance of potentially slippery sections of road.

Pavement Structural Assessment

Flexible pavement structures usually deteriorate under traffic loading over time, as evidenced by structural defects (e.g. rutting and fatigue cracking in the wheel track), which affects the structural integrity. Therefore, the aim of structural assessment is to identify the ability of a pavement to carry the anticipated future traffic load in a design period, which will usually determine the remaining life of the pavement (i.e. the cumulative number of standard axle loads) at a predefined pavement condition. The result of such assessment are also used to determine remedial or strengthening works (e.g. overlay or rehabilitation needs) (Flaherty, 2001).

Structural performance is usually determined by a deflection survey. As a quantitative indicator, the magnitude of deflection under a known load can be associated with the strength of the pavement layers and subgrade, which could be a reasonable estimate of pavement remaining life (i.e. determining weak structure) and, if required, the overlay thickness (i.e. through back analysis procedure). The higher the measured deflection indicates the worse structure.

A number of devices have been developed to measure the deflection of pavement under static (i.e. falling weight deflectometer — FWD) or impulse (e.g. high-speed defectograph) loading (cf. [Section 2.3.2](#)). While these are widely used devices for structural assessment, their relative disadvantages, such as labour-intensive point measurements for FWD and the high cost of defectographs, present difficulties for large-scale road network measurement. Alternatively, if relatively low-cost smartphone applications could be used to determine pavement deflection under vehicle loading (i.e. mounted in a heavy truck), this would be beneficial to road agencies by reducing costs.

Theoretically, when a smartphone is mounted in a heavy moving truck, deflection of the

pavement could be included in the measured vehicle vertical acceleration. In practice, the amount of measured deflections that are reflected in a vehicle's vertical acceleration are too small to be recognised in measured accelerations that contain noises caused by the vehicle's dynamics; however, research has found that higher deflection may be generally associated with wheel rutting or areas of cracking (Flaherty, 2001). Although this may not always be the case, it would be a potentially fruitful to undertake a spectral analysis of vehicle vertical body accelerations (Loprencipe and Cantisani, 2013). Because a road profile can be decomposed into a series of sinusoids, with each having a different wavelength, amplitude and phase (see [Section 2.3.4](#)), short wavelength components (less than 1 m) are associated with road surface textures affecting road friction and tyre wear. A medium wavelength (1 to 20 m) refers to surface deterioration and structural deformation while long wavelengths are associated with movements in the pavement foundation.

A focus of attention could be to consider medium and long wavelengths in a spectral analysis of vertical acceleration, particularly for wavelengths greater than 5 m (i.e. wavenumber of less than 5^{-1} cycle/m). Some examples of a typical acceleration spectrum, measured by a known vehicle travelling at a certain speed on a known road profile, are required to show the upper limits/trigger value of the acceleration spectrum in which the structural condition of such a section is unacceptable (e.g. where rut depth and surface deformation is high). Thereafter, it is possible to determine whether or not the measured road section has more serious structural defects than those example road sections.

Section 9.5 Unpredictable Factors Affecting Roughness Measurement

Chapter 5 reported on an investigation of the effects of a number of factors on the measurement of roughness associated with the smartphone itself, vehicle characteristics

and the road profile. In practice, there are some other conditions that can occur, the effects of which are difficult to predict.

These factors include perturbations in the road surface, such as potholes and speed bumps, which can cause a large variance in measured vehicle body acceleration, depending on vehicle speed and vehicle-related factors, such as low vehicle operating speed on the roads that highlight many defects, rapid acceleration/deceleration, vehicle cornering, and smartphones suddenly falls. Such factors can occur at any time during the measurement of road conditions using smartphone-based applications; however, their effects on vehicle body acceleration are relatively difficult to quantify and correct for due to their randomness. To reduce the effects of these factors on the accurate measurement of roughness, filters can be used to limit both the amplitude and variability of measured acceleration data.

A very basic band-pass filter, that only records the vertical acceleration readings within a typical range (e.g. +1 to -1g) could be considered in order to exclude the effect of extreme movement of the smartphone (e.g. smartphone falling or a step input from a pothole or speed bump). Carrera, Guerin and Thorp (2013) adopted such a process to identify potholes. The second type of filter to consider would be one for vehicle speed, which could detect low-speed situations (e.g. parking and moving off from a stop), in order to halt the measurement when the vehicle is driven at low speeds (e.g. lower than 15 m/s).

Section 6.3.6 contained the results from an investigation into the impact of driver behaviour on vertical vehicle acceleration, where in it was found that the effects of normal acceleration and deceleration are insignificant. This study did not consider aggressive driving, however, including rapid acceleration and braking. Such driving patterns may

affect the measurement of road roughness by causing variance in the measured vehicle body acceleration. To address this, longitudinal acceleration filters can be adopted, which do not record data when vehicle longitudinal acceleration or deceleration is outside of a given range (e.g. + 0.5 to -0.5 g) (Byrne *et al.*, 2013).

Potholes

Potholes are considered to be the most common road defect by many road users; however, people are seldom aware that potholes are, in fact, relatively expensive to repair. For example, in England during 2017, the total number of potholes filled was 1.7 million, with a pothole being filled every 21 seconds, and the total cost of pothole patching estimated at £102 million. Furthermore, over the last decade, the total expenditure for pothole patching was found to be £1 billion (Asphalt Industry Alliance, 2017). Whilst potholes result in further serious damage to the road surface if they are not repaired, they can also cause significant damage to the road user. In England, in 2017, compensation claims reached average of 5 million per authority, of which 70% were pothole-related.

The occurrence of potholes is difficult to predict, in terms of where and when they might appear. To address this, a number of smartphone-based approaches have been used to detect potholes (Carrera *et al.*, 2013; Mednis *et al.*, 2011; Eriksson *et al.*, 2008). Approaches adopted include the use of a dynamic threshold determined from the variance of acceleration data to identify potential potholes (Byrne *et al.*, 2013); and a simple static threshold allied to crowdsourced repeat confirmations by multiple users (Carrera *et al.*, 2013). Such approaches could easily be included within the system advocated here to measure both road roughness and detect potholes.

Speed Bumps

Traffic calming measures are being introduced on local road networks to reduce vehicle speed and improve road safety, particularly in residential areas. A typical example is the speed bump, which is designed to reduce vehicle speed. Considering the smartphone-based application, during data acquisition, road bumps can cause a large variance in the measured vertical vehicle acceleration. As a result, the perceived roughness for a vehicle can increase (T. D. Gillespie, 1992b). **Figure 9.2** shows such an example in which the effect of a long concrete speed bump on measured vehicle body acceleration is simulated. **Figure 9.2** demonstrates that measured vehicle body accelerations did not simply follow the shape of the speed bump, but rather showed large variances of measured acceleration on the uphill and downhill slopes. Such effects are amplified with increased vehicle speed.

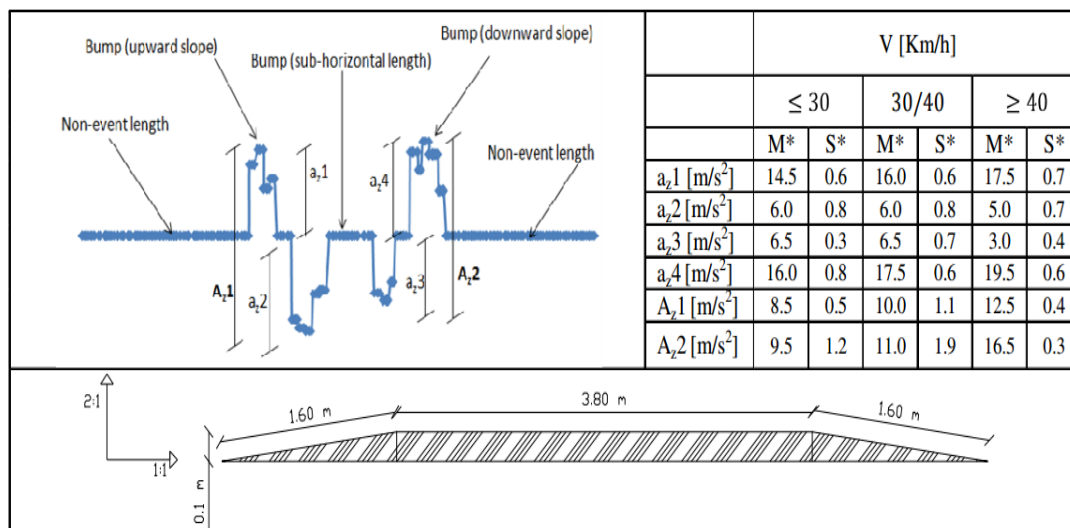


Figure 9.2: Vehicle vertical acceleration peak values for concrete bump (M^* =average value, S^* =standard deviation) (Astarita *et al.*, 2012)

In order to understand this phenomenon better, further simulations were carried out using CarSimTM. A real speed bump (Weber, 1998) was embedded into a section of road 150 m

in length (see **Figure 9.3**).

The original IRI of this section (120 m in length) is 2.4 m/km, whereas after modification, the IRI increased to 4.4 m/km.

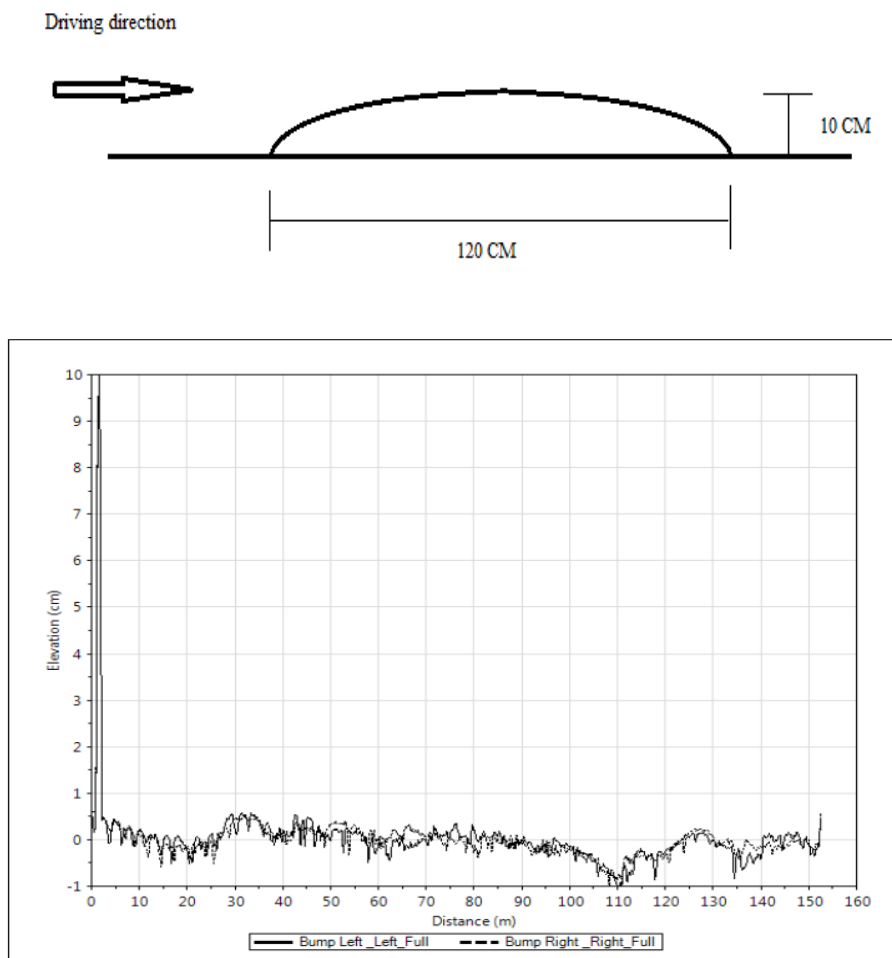
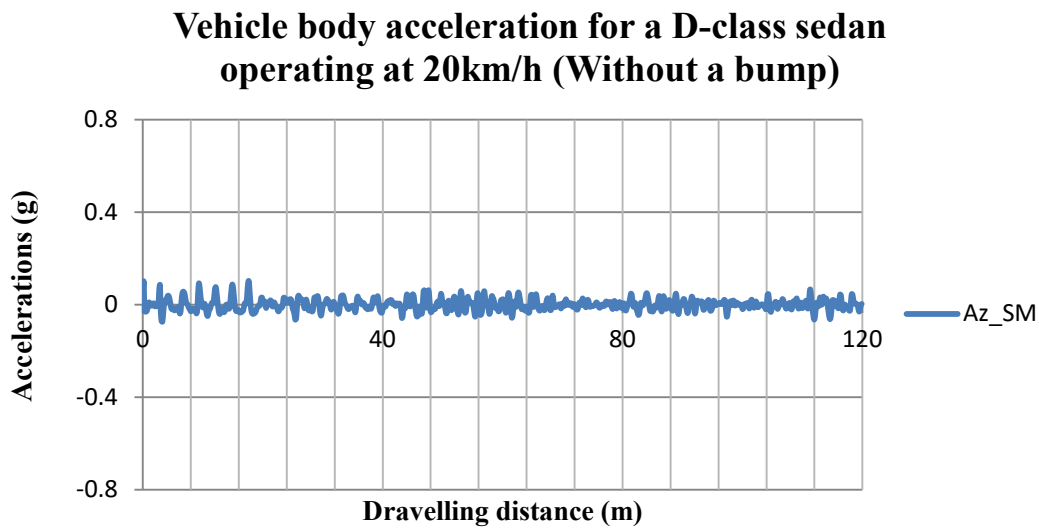
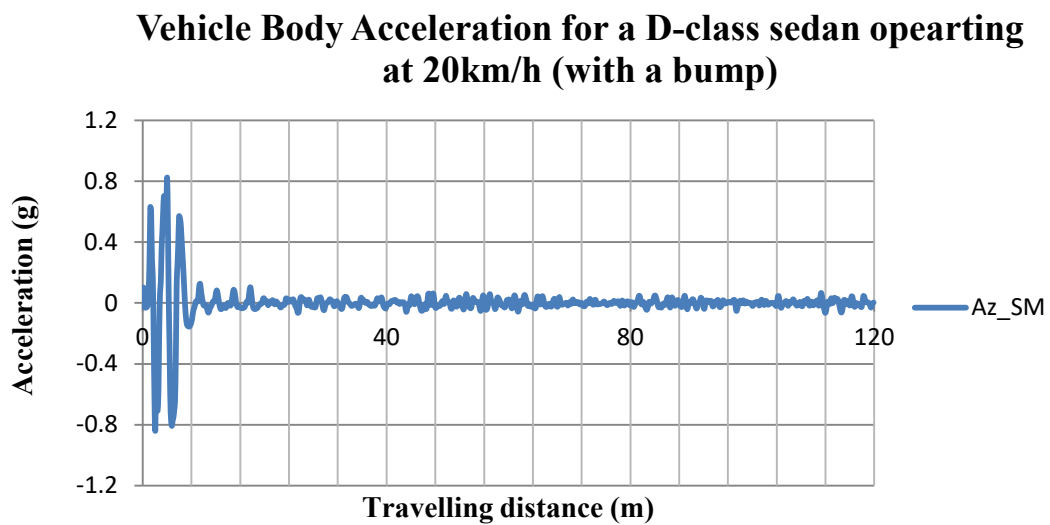


Figure 9.3: Road profile modeified (with a bump), IRI=4.4 m/km

The vehicle chosen for the simulation was a D-class sedan, travelling at a speed of 20km/h, with resulting measured vehicle body accelerations shown in **Figure 9.4** and **Figure 9.5**.



*Figure 9.4: Vehicle body acceleration measured on original road profile
(IRI=2.4/km)*



*Figure 9.5: Vehicle body acceleration measured on modified road profile
(IRI=4.4m/km)*

After calculation, the simulated Grms of the original section (without a bump) was 0.024 g, whereas the Grms calculated for the section with a bump was 0.104 g. For comparison, the simulated Grms for another road section with a similar profile, and with a similar IRI value of 4.4 m/km (with a bump), was calculated as 0.052 g (using the same test vehicle running

at the same speed). From the results, it can be seen that adding a bump to the road profile can significantly increase the measured vehicle body accelerations (i.e. an increase of 200% from 0.052 to 0.104 g in the above example). Evidently, the imposition of a speed bump is not a reflection of the true condition of the road, and its effect on measured road roughness might not be of interest to a road agency concerned with using roughness to identify maintenance issues; however, it may be of interest to the road user, as its effect on road use costs may be significant.

Section 9.6 Use of smartphone data to support road asset management

Data from the developed prototype system can be used to support asset management decisions. Appendix B shows how IQL-3/4 data obtained from smartphone-based applications can be used within a decision support tool (i.e. HDM-4) to determine economic road maintenance standards. Furthermore, although it was not demonstrated in the research, continuously updated data are also useful for road network condition monitoring allowing the condition of a road network to be screened frequently and thereby enabling road sections requiring further detailed maintenance treatments to be identified. Not only can road agencies benefit from the developed system in this way, but road users could also be provided with benefits. For example, the system proposed here in could enable the road user to be provided with the condition of the road they will be travelling on, their road use costs per mile (i.e. fuel costs, vehicle damage/depreciation costs and delay costs, etc.) thus enabling them to make informed decisions about their journeys.

Section 9.7 Benefits of the Smartphone-Based System

Encouraging the use of smartphone-based applications to measure road conditions would transfer resource requirements from road agencies to road users, saving road agencies time and expense. Clearly, road users would eventually benefit from using such systems, although indirect benefits due to improved road conditions and decreased road use costs might not be immediately apparent to road users, as such changes are generally intangible and small.

Today, the use of smartphone-based applications in vehicles is very common due to convenience and simplicity of use (e.g. GPS navigation). According to the literature, the number of smartphone users in the UK was 42 million in 2017, representing approximately 63% of the total population (eMarketer, 2017). This is likely to increase rapidly over time, as younger generations grow up with smartphone technology as part of their daily lives. Hence, most people in the UK today have access to smartphones, and are conversant in their use and associated applications (apps). With embedded applications in smartphones running in the background to automatically capture vehicle body acceleration, most road users would not need to spend any additional time in achieving such a function.

Motivating the public to use a smartphone prototype system for measuring road conditions would be beneficial. This can be considered from two different perspectives: a) to encourage companies, such as Uber TM and Waze TM, and logistics companies to cooperate with road agencies to provide road condition data; and b) to encourage the public to directly use smartphone-based applications on a daily basis to provide road condition data. This is because, in the UK, the total road length is estimated to be 246,500 miles (0.39 million km). Of this length, 80% (196,300 miles) is represented by minor roads (i.e. Classes C and U roads) in both urban and rural areas (Department for Transport, 2017).

If, for example, in the UK, the average cost of measuring road conditions is £5 per mile (McGhee, 2004), the total expenditure for road agencies to measure road conditions for C and U class roads only is in an order of £1 million annually. However, significant savings to the road agency (and tax payer) could accrue if the data collection exercise was carried out, for an agreed fee, by taxi companies and/or logistics companies that already traverse large parts of the road network. Many of the personnel working for such companies are already likely to be smartphone technology savvy and, therefore, the addition of an app to those they already use would be relatively straightforward.

In addition, many road users already use apps and other means to report the presence of discrete road defects, such as potholes (Carrera *et al.*, 2013). Indeed, local authorities and road maintainers provide web-based means by which the public can report road defects (Birmingham City Council, 2018). It could therefore be envisioned that certain members of the public might voluntarily use a freely available app for measuring road condition. Furthermore, the public might be further encouraged if their measured road condition data is made publicly available, together with the associated road use costs. By so doing, local authorities would be encouraged to pay more attention to maintaining their local road networks. By linking road use costs to actual routes, the public would be provided with an additional source of information with which to make informed decisions regarding their road networks. Moreover, local authorities might be shamed into maintaining those road sections which have significantly higher total road use costs compared to other routes.

Section 9.8 Additional Considerations

There are other aspects that may also draw our attention to the development and implementation of a prototype system. Today, the built-in accelerometer in vehicles has

been a common device included in the manufacturing of a vehicle, particularly for electric vehicles. It is argued that the use of these mounted accelerometers in a moving vehicle can provide information on vehicle dynamics (e.g. to capture vehicle body acceleration) in the same way as the smartphone application does; however, it may be found that these accelerometers are not cloud-based, so internet connections to enable data to be sent to a cloud server for further analysis would be needed. To enable such connections, more costs and time would be incurred on the manufacturing side, as current data captured by accelerometers are sent to an in-vehicle computer to monitor vehicle performance.

Crowdsourcing is a key concept in the development of a smartphone prototype system that generally requests multiple confirmations from road users in the identification of data measured. Not only could this concept be used in automated road condition measurement using a smartphone application (e.g. roughness and potholes), but this concept could also be extended to visual inspection of road conditions (e.g. cracking) or other incidents (i.e. flooding, accident reports, etc.).

In addition, in this study, simulated data were assumed to be equal to smartphone captured data in order to demonstrate the prototype system in theory, whereas practical laboratory work is not presented. When real measurements are included in the system for analysis, the level of accuracy may be reduced; however, this will not reduce the usefulness of the theoretical system, because the accuracy can be improved by other means, whereas the concept behind the system does not change.

Section 9.9 Summary

In this chapter, the fulfilment of the predefined objectives, and other aspects that were not considered in the development and implementation of the prototype system, have been

discussed. Appropriate suggestions have been made to enable improvements to the system to become more effective and efficient in future work. It was recognised that, although this prototype system is in an early stage of development and requires further refinement, the provisional results have indicated that it may provide an effective measurement of road roughness (IQL-2), considering the predominant factors. It was also found that the random forest algorithm would be suitable for the analysis of road conditions captured from various vehicles. Data from the developed prototype system could be used to support asset management decisions, including determining economic maintenance standards and estimating the economic benefits derived from selected standards.

Apart from measuring road roughness, the prototype system developed may also be used to measure other data required to support management decisions, such as skid resistance (related to road safety) and possibly structural conditions. An additional technique, the use of filters, was also suggested for enhancing the quality of the data measured, particularly in dealing with unpredictable situations, such as low vehicle operating speeds on roads that highlight many defects (e.g. potholes), rapid acceleration/deceleration, vehicle cornering and smartphones suddenly fall. The problem of potholes and speed bumps was also discussed in detail, with suggestions for dealing with these factors. Furthermore, this section presented ideas that would encourage both commercial and public use of the system by showing how they could benefit through reducing expenditure on data collection and providing better road conditions.

Chapter 10

CONCLUSION AND FURTHER RESEARCH

Section 10.1 Accomplished work

The work presented in this thesis has demonstrated the development of a prototype system which uses road condition data collected using smartphone technology mounted within moving vehicles to support road asset management decision making. The research has focused on the measurement of road roughness, which is a widely used measure of both road condition and vehicle operating costs. The research conducted to develop the prototype system comprised of the following processes:

1. The identification of the data required for the assessment of road condition in order to support road maintenance decisions.
2. An exploration of the potential for a smartphone fitted to a moving vehicle to collect the above identified measures of road condition via literature review and practical application.
3. The identification and quantification of the predominant factors that affect the measurement of roughness using a smartphone mounted within a moving vehicle including:
 - (i) Smartphone-related factors
 - (ii) External factors including those to do with the vehicle and the road profile
 - (iii) Development and use of a computational model to ascertain roughness from vehicle vertical acceleration taking into account the identified predominant factors

4. Building and testing a prototype system that could utilize data to determine measures of road roughness. This involved:
 - (i) The use of a vehicle dynamics package, CarSimTM, to generate a dataset consisting of 0.8 million data points
 - (ii) Consideration, testing and selection of an appropriate machine learning algorithm
 - (iii) Determining the level of accuracy that the system can achieve as a function of the availability of information regarding the variables which were found to influence the measurement of vehicle vertical acceleration.
5. Demonstrating the usefulness of the data provided by the prototype system within an existing road asset management system.

Section 10.1.1 Conclusion

The conclusions from the research are as follows:

1. The system can use data collected using smartphone technology fitted within moving vehicles to measure road roughness.
2. The system is capable of obtaining precise measures of road roughness (i.e. IQL-2) as well as classifying road condition (i.e. IQL-3/4) depending on the influencing variables taken into account.
3. Further investigations are required in order for a similar system to that demonstrated here to be developed to measure structural road condition and skid resistance.

Section 10.1.2 Findings

A number of findings were identified as follows.

Section 10.1.2.1 The smartphone application

1. When securely fastened to a moving vehicle, a smartphone with a 3-axis accelerometer should theoretically be able to measure road roughness and other major defects that cause vehicle vibration, as well as providing useful information on road surface friction.
2. Converting the vehicle body acceleration to root-mean-square (RMS) vehicle body acceleration (Grms) can improve the prediction of IRI
3. The parametric analysis found the vehicle type, speed and the peculiar characteristics of the smartphone influenced most the measured vehicle body acceleration and therefore the measurement of roughness.
4. A data collection system utilizing a smartphone can satisfy the frequency of data collection (with sampling interval of 300 mm or less) and the accuracy requirements in the ASTM international standard as a Class 3 device.
5. Machine learning algorithms performed better than multivariable linear regression analysis. The multivariable linear regression analysis techniques trialled were found to be able to predict IRI to an accuracy that is equivalent to a visual inspection (i.e., Class 4 device at IQL-4).

Section 10.1.2.2 Develop a prototype system

1. A random forest algorithm was identified as the most appropriate machine-learning algorithm for both predicting the IRI accurately and also for classifying road condition in terms of ranges of IRI.
2. The prototype system using simulated data can measure IRI to within an error of

± 0.3 m/km when vehicle speed, type, Grms and sprung mass data and smartphone types are available. The results presented suggested that the computational model is able to predict IRI to an accuracy that is suitable for IQL-2 measurement.

3. The errors in predicting IRI from a simulated dataset containing multiple vehicle types was related to the proportion of commercial vehicles (i.e. Utility Truck (UT), Compact Pickup (CP), Large European Van (LEV), European Van (EV) and D-class Minivan respectively).

Section 10.1.2.3 usefulness of data generated from prototype system

1. Network-level data can be used to set economic maintenance standards and estimate associated road agency and road user expenditure via a strategic analysis using the World Bank's road investment appraisal model, HDM-4 (see Appendix B).
2. Taking into account possible errors introduced by the smartphones themselves, may result in a decrease in the accuracy of the measurement of road roughness (e.g. from IQL-2 to IQL-3). Although it is possible to rebuild the algorithm to achieve higher accuracy, it is balanced between the complexity of the algorithm and the desired accuracy of the system.
3. The methodology used in the development of machine learning can provide fundamental understanding in building real world applications.
4. Outputs from the above system can be widely used in developing countries where circumstances prohibit the use of data collection systems which are expensive to operate and maintain. The outputs are also useful in many developed worlds as the initial diagnosis of the network condition that informs further field

measurement.

Section 10.1.3 Remaining problems

Section 10.1.3.1 The use of smartphone applications

Apart from measuring road roughness a comprehensive assessment of road condition requires the following additional information to be measured:

- a) Pavement strength (to determine the structural capacity of a road and thereby its residual life)
- b) Skid resistance (a measure of road safety)

Measuring pavement strength

The following findings with respect to the measurement of pavement strength from data captured by the smartphone application fitted in a moving vehicle have been identified.

1. Theoretically, it should be possible to measure the deflection of a road pavement caused by the passage of a heavy truck using a smartphone mounted inside the heavy truck. In practice, the measured deflections caused by the vehicle and therefore captured as the vehicle vertical acceleration are likely to be so small that they cannot be isolated from noise, such as the measured accelerations caused by vehicle dynamics.
2. To address this it might be possible to utilize an IQL-3/4 approach whereby in vehicle acceleration spectrum, to determine limits/ trigger value in the vicinity of particular wavelengths where the road condition of such a section is unacceptable.

Measuring road friction

It was postulated that smartphone application may be used to identify roads of unacceptably low skid resistance, albeit considerable research may be required. Possible avenues of research include:

1. Developing a correlation between the average stopping distances of vehicles undergoing an emergency stop (with known initial vehicle speed) with road skid resistance in wet condition.
2. The examination of the measured vehicle deceleration values (e.g. at junctions) to determine whether a driver is performing an emergency stop.
3. The use of multiple confirmations by users to determine road sections with insufficient levels of skid resistance particularly at traffic junctions, approach lanes (i.e., arms) to the roundabout and pedestrian crossings i.e. locations where drivers are likely to perform an emergency stop.

Section 10.1.3.2 Additional considerations

To enable a system successfully to capture and analyse the vehicle body acceleration that caused by road roughness and other structural defects, it was suggested to have the following additional features:

1. Multiple filters to detect the effect of road profile related defects (i.e. pothole and speed bumps), low vehicle speed, hard accelerating/decelerating of sport cars, vehicle cornering and the sudden falling of the smartphone.
2. Strategies should be introduced from the point of view of the road agency to encourage public or commercial use of smartphone application to measure road condition. Both road agency and users could be benefited from the proposed prototype system.

Section 10.2 Future work

To enable a fully working system to be developed for the measurement of road condition using a smartphone application, further works concerning data collection and analysis are needed, including:

- a) The systematic laboratory testing and comparison of various smartphone models enabling smartphones to be classified on their relative performance according to a number of metrics; and
- b) Comparing vehicle vertical accelerations measured with a smartphone and other accelerometers with that predicted by CarSimTM. A possible approach might be to compare actual accelerators fitted inside a moving vehicle with the data obtained from a smartphone and also to simulate same experiment in CarSimTM.
- c) Calibration of the prototype system by comparing measured roughness with that predicted by the proposed system.

REFERENCES

- Abdelaziz, N., Abd El-Hakim, R.T., El-Badawy, S.M. and Afify, H.A. (2018) 'International Roughness Index prediction model for flexible pavements', *International Journal of Pavement Engineering*, , pp. 1-12.
- Abdel-Hamid, W. (2004) *Accuracy enhancement of integrated MEMS-IMU/GPS systems for land vehicular navigation applications*. University of Calgary, Department of Geomatics Engineering.
- Ahmed, M. and Abdel-Aty, M. (2013) 'Application of Stochastic Gradient Boosting Technique to Enhance Reliability of Real-Time Risk Assessment: Use of Automatic Vehicle Identification and Remote Traffic Microwave Sensor Data', *Transportation Research Record: Journal of the Transportation Research Board*, (2386), pp. 26-34.
- Aki, M., Rojanaarpa, T., Nakano, K., Suda, Y., Takasuka, N., Isogai, T. and Kawai, T. (2016) 'Road Surface Recognition Using Laser Radar for Automatic Platooning.', *IEEE Trans.Intelligent Transportation Systems*, 17(10), pp. 2800-2810.
- Alexopoulos, E.C. (2010) 'Introduction to multivariate regression analysis', *Hippokratia*, 14(Suppl 1), pp. 23-28.
- Aljeroudi, Y., Legowo, A. and Sulaeman, E. (2006) 'MOBILITY DETERMINATION AND ESTIMATION BASED ON SMARTPHONES-REVIEW OF SENSING AND SYSTEMS', .
- Analog Devices (2018) *Accelerometers*. Available at: <http://www.analog.com/en/parametricsearch/11175> (Accessed: 07/08/2018).
- Archondo-Callao, R. (2009) *RONET Road Network Evaluation Tools, User Guide, Version 2.0*, .
- Asphalt Industry Alliance (2017) *Annual Local Authority Road Maintenance Survey*. UK: Available at:(Accessed: .
- Astarita, V., Caruso, M.V., Danieli, G., Festa, D.C., Giofrè, V.P., Iuele, T. and Vaiana, R. (2012) 'A mobile application for road surface quality control: UNIquALroad', *Procedia-Social and Behavioral Sciences*, 54, pp. 1135-1144.
- ASTM, E. (2015) '1926-08', "Standard Practice for Computing International Roughness Index of Roads from Longitudinal Profile Measurements", *ASTM International*, .
- ASTM, E. (2012) '1082-90', "Standard Test Method for Measurement of Vehicular Response to Traveled Surface Roughness", *ASTM International*, .

ASTM, E. (2009) '950', *"Standard Test Method for Measuring the Longitudinal Profile of Traveled Surfaces with an Accelerometer Established Inertial Profiling Reference"*, ASTM International, .

Attoh-Okine, N.O. (1999) 'Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance', *Advances in Engineering Software*, 30(4), pp. 291-302.

Attoh-Okine, N.O. (1994) *Predicting roughness progression in flexible pavements using artificial neural networks*.

Bailey, D. and Thompson, D. (1990) 'How to develop neural-network applications', *AI expert*, 5(6), pp. 38-47.

Belzowski, B. and Ekstrom, A. (2015) 'Evaluating roadway surface rating technologies', .

Benekohal, R. and Treiterer, J. (1988) 'CARSIM: Car-following model for simulation of traffic in normal and stop-and-go conditions', *Transportation Research Record*, (1194).

Bennett, C.R. and Paterson, W.D. (2000) 'A guide to calibration and adaptation', *HDM-4. Volume 5. The Highway Development and Management Series*, .

Bennett, C.R., Solminihaq, H.d., Chamorro, A., Chen, C. and Flintsch, G.W. (2007) *Data Collection Technologies for Road Management*. Washington, D.C: East Asia Pacific Transport Unit, World Bank. Available at: (Accessed: .

Bennett, C. and Greenwood, I. (2003) 'Volume 5: HDM-4 calibration reference manual', *International Study of Highway Development and Management Tools (ISOHDM)*, .

Berk, R.A. (2005) 'Data mining within a regression framework' *Data Mining and Knowledge Discovery Handbook* Springer, pp. 231-255.

Bhavsar, H. and Ganatra, A. (2012) 'A comparative study of training algorithms for supervised machine learning', *International Journal of Soft Computing and Engineering (IJSCE)*, 2(4), pp. 2231-2307.

Biau, G. (2012) 'Analysis of a random forests model', *Journal of Machine Learning Research*, 13(Apr), pp. 1063-1095.

Birmingham City Council (2018) *Report a road or pavement problem*. Available at: https://www.birmingham.gov.uk/info/20110/report_road_and_pavement_issues/643/report_a_road_or_pavement_problem (Accessed: 17/10/2018).

Bradbury, T., Parkman, C., Abell, R. and Johnston, K. (2012) *What Are the Wider Impacts of Changes to Road Maintenance Budgets?* .

Breiman, L. (2001) 'Random forests', *Machine Learning*, 45(1), pp. 5-32.

- Breiman, L. (2000) 'Some infinity theory for predictor ensembles', .
- Breiman, L. (1996) 'Bagging predictors', *Machine Learning*, 24(2), pp. 123-140.
- Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A. (1984) *Classification and regression trees*. CRC press.
- Bridgelall, R. (2013) 'Connected vehicle approach for pavement roughness evaluation', *Journal of Infrastructure Systems*, 20(1), pp. 04013001.
- Briem, G.J., Benediktsson, J.A. and Sveinsson, J.R. (2002) 'Multiple classifiers applied to multisource remote sensing data', *IEEE Transactions on Geoscience and Remote Sensing*, 40(10), pp. 2291-2299.
- Burrow, M., Evdorides, H., Wehbi, M. and Savva, M. (2013) 'The benefits of sustainable road management: a case study', *Proceedings of the Institution of Civil Engineers, Transport*, 166, pp. 222-232.
- Buttlar, W.G. and Islam, M.S. (2014) *Integration of Smart-Phone-Based Pavement Roughness Data Collection Tool with Asset Management System*, .
- Byrne, M., Parry, T., Isola, R. and Dawson, A. (2013) 'Identifying road defect information from smartphones', .
- Cai, P., Wang, Y., Lu, G. and Chen, P. (2015) *An Improved K-Nearest Neighbor Model for Road Speed Forecast Based on Spatiotemporal Correlation*.
- Cantisani, G. and Loprencipe, G. (2010) 'Road roughness and whole body vibration: Evaluation tools and comfort limits', *Journal of Transportation Engineering*, 136(9), pp. 818-826.
- Carrera, F., Guerin, S. and Thorp, J. (2013) 'By the people, for the people: the crowdsourcing of “streetbump”: an automatic pothole mapping app', *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-4/W1*, , pp. 19-23.
- Caruana, R. and Niculescu-Mizil, A. (2006) *An empirical comparison of supervised learning algorithms*. ACM, pp. 161.
- Chai, T. and Draxler, R.R. (2014) 'Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature', *Geoscientific model development*, 7(3), pp. 1247-1250.
- Chan, J.C., Huang, C. and Defries, R. (2001) 'Enhanced algorithm performance for land cover classification from remotely sensed data using bagging and boosting', *IEEE Transactions on Geoscience and Remote Sensing*, 39(3), pp. 693-695.

- Chandra, S., Sekhar, C.R., Bharti, A.K. and Kangadurai, B. (2012) 'Relationship between Pavement Roughness and Distress Parameters for Indian Highways', *Journal of Transportation Engineering*, 139(5), pp. 467-475.
- Chang, L. and Chen, W. (2005) 'Data mining of tree-based models to analyze freeway accident frequency', *Journal of Safety Research*, 36(4), pp. 365-375.
- Chang, L. and Wang, H. (2006) 'Analysis of traffic injury severity: An application of non-parametric classification tree techniques', *Accident Analysis & Prevention*, 38(5), pp. 1019-1027.
- Chen, K., Lu, M., Fan, X., Wei, M. and Wu, J. (2011) *Road condition monitoring using on-board Three-axis Accelerometer and GPS Sensor*. IEEE, pp. 1032.
- Chipman, H.A., George, E.I. and McCulloch, R.E. (1998) 'Bayesian CART model search', *Journal of the American Statistical Association*, 93(443), pp. 935-948.
- Chou, P.A. (1991) 'Optimal partitioning for classification and regression trees', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4), pp. 340-354.
- Chugh, G., Bansal, D. and Sofat, S. (2014) 'Road Condition Detection Using Smartphone Sensors: A Survey', *International Journal of Electronic and Electrical Engineering*, 7(6), pp. 595-601.
- Costello, S.B., Snaith, M.S., Kerali, H., Tachtsi, L.V. and Ortiz-García, J.J. (2005) *Stochastic model for strategic assessment of road maintenance*. Thomas Telford Ltd, pp. 203.
- Cunningham, P. and Delany, S.J. (2007) 'k-Nearest neighbour classifiers', *Multiple Classifier Systems*, , pp. 1-17.
- Danjuma, K.J. (2015) 'Performance Evaluation of Machine Learning Algorithms in Post-operative Life Expectancy in the Lung Cancer Patients', *arXiv preprint arXiv:1504.04646*, .
- Dasgupta, A., Sun, Y.V., König, I.R., Bailey-Wilson, J.E. and Malley, J.D. (2011) 'Brief review of regression-based and machine learning methods in genetic epidemiology: the Genetic Analysis Workshop 17 experience', *Genetic epidemiology*, 35(S1), pp. S5-S11.
- Davis, B.R. and Thompson, A. (2001) 'Power spectral density of road profiles', *Vehicle System Dynamics*, 35(6), pp. 409-415.
- Dawkins, J., Bishop, R., Powell, B. and Bevly, D. (2011) 'Investigation of Pavement Maintenance Applications of Intellidrive SM (Final Report): Implementation and Deployment Factors for Vehicle Probe-based Pavement Maintenance (PBPM)', *Auburn University*, .

De'ath, G. and Fabricius, K.E. (2000) 'Classification and regression trees: a powerful yet simple technique for ecological data analysis', *Ecology*, 81(11), pp. 3178-3192.

Del Rosario, M.B., Redmond, S.J. and Lovell, N.H. (2015) 'Tracking the evolution of smartphone sensing for monitoring human movement', *Sensors*, 15(8), pp. 18901-18933.

Demuth, H. and Beale, M. (2001) 'Neural network toolbox user guide (Version 4). 3 Apple Hill Drive Natick 01760-2098. The Math-Works', *Inc., Massachusetts*, .

Department for Transport (2017) *Road Lengths in Great Britain 2016*, GOV. UK. UK: GOV.UK. Available at: <https://www.gov.uk/government/statistics/road-lengths-in-great-britain-2016> (Accessed: 2018/09/12).

Dodds, C. and Robson, J. (1973) 'The description of road surface roughness', *Journal of Sound and Vibration*, 31(2), pp. 175-183.

Doertenbach, N. (2012) *The Calculation of Grms*. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.168.3159&rep=rep1&type=pdf> (Accessed: .

Doré, G., Flamand, M. and Pascale, P. (2002) 'Analysis of the wavelength content of the longitudinal profiles for C-LTPP test sections', *Canadian Journal of Civil Engineering*, 29(1), pp. 50-57.

Doreswamy, H.K. (2012) 'Performance Evaluation of Predictive Classifiers for Knowledge Discovery from Engineering Materials Data Sets', *arXiv preprint arXiv:1209.2501*, .

Douangphachanh, V. and Oneyama, H. (2014) *Formulation of a simple model to estimate road surface roughness condition from Android smartphone sensors*. IEEE, pp. 1.

Douangphachanh, V. and Oneyama, H. (2013) 'A study on the use of smartphones for road roughness condition estimation', *Journal of the Eastern Asia Society for Transportation Studies*, 10(0), pp. 1551-1564.

Dreiseitl, S. and Ohno-Machado, L. (2002) 'Logistic regression and artificial neural network classification models: a methodology review', *Journal of Biomedical Informatics*, 35(5), pp. 352-359.

Du, Y., Liu, C., Wu, D. and Jiang, S. (2014) 'Measurement of International Roughness Index by Using Z-axis Accelerometers and GPS', *Mathematical Problems in Engineering*, 2014.

DuPont, E.M., Roberts, R.G., Selekwa, M.F., Moore, C.A. and Collins, E.G. (2005) *Online terrain classification for mobile robots*. American Society of Mechanical Engineers, pp. 1643.

Dupuy, S., Egges, A., Legendre, V. and Nugues, P. (2001) *Generating a 3D simulation of a car accident from a written description in natural language: The Carsim system*. Association for Computational Linguistics, pp. 1.

Durst, P.J., Mason, G.L., McKinley, B. and Baylot, A. (2011) 'Predicting RMS surface roughness using fractal dimension and PSD parameters', *Journal of Terramechanics*, 48(2), pp. 105-111.

El-Diasty, M. and Pagiatakis, S. (2008) 'Calibration and stochastic modelling of inertial navigation sensor errors', *Journal of Global Positioning Systems*, 7(2), pp. 170-182.

Elkan, C. (2013) 'Predictive analytics and data mining', Online: <http://cseweb.ucsd.edu/~elkan/255/dm.pdf>, .

ElMoustapha, O., James, W., Charles, Y. and Kshitij, A. (2007) *On the Comparison of Regression Algorithms for Computer Architecture Performance Analysis of Software Applications*.

eMarketer (2017) *UK Digital Users: The eMarketer Forecast for 2017*. Available at: <https://www.emarketer.com/Report/UK-Digital-Users-eMarketer-Forecast-2017/2001988> (Accessed: 27/08/2018).

Emerson, D., Nayak, R. and Weligamage, J. (2011) 'Using data mining to predict road crash count with a focus on skid resistance values', .

Emery, W. and Singh, M.C. (2013) *Large-Area Road-Surface Quality and Land-Cover Classification Using Very-High Spatial Resolution Aerial and Satellite Data*. Remote Sensing of Roads and Highways in Colorado. Available at:(Accessed: .

Eriksson, J., Girod, L., Hull, B., Newton, R., Madden, S. and Balakrishnan, H. (2008) *The pothole patrol: using a mobile sensor network for road surface monitoring*. ACM, pp. 29.

Federal Highway Administration and Association of State Highway and Transportation Officials (1996) 'Asset management advancing the state of the art into the 21st century through public-private dialogue', .

Feng, M., Fukuda, Y., Mizuta, M. and Ozer, E. (2015) 'Citizen Sensors for SHM: Use of Accelerometer Data from Smartphones', *Sensors*, 15(2), pp. 2980-2998.

Ferligoj, A. and Kramberger, A. (1995) *Some Properties of R^2 in Ordinary Least Squares Regression*. Fakulteta za družbene vede, pp. 133.

Fernández, C., Huerta, I. and Prati, A. (2015) 'A comparative evaluation of regression learning algorithms for facial age estimation'*Face and Facial Expression Recognition from Real World Videos* Springer, pp. 133-144.

FHWA (2018) *infopave*. Available at: <https://infopave.fhwa.dot.gov/> (Accessed: .

- Flaherty, C.A. (2001) *Highways: The Location, Design, Construction & Maintenance of Road Pavements*. Butterworth-Heinemann.
- Forslöf, L. (2012) *Roadroid–Smartphone Road Quality Monitoring*. Austria: .
- Freund, Y. and Schapire, R.E. (1996) *Experiments with a new boosting algorithm*. pp. 148.
- Friedman, J.H. (2001) 'Greedy function approximation: a gradient boosting machine', *Annals of statistics*, , pp. 1189-1232.
- Friedman, J.H. and Hall, P. (2007) 'On bagging and nonlinear estimation', *Journal of statistical planning and inference*, 137(3), pp. 669-683.
- Gao, J. and Chen, K. (2011) 'Frequency-Domain Simulation and Analysis of Vehicle Ride Comfort based on Virtual Proving Ground', *Int J Intell Eng Syst*, 4(3), pp. 1-8.
- García-Pedrajas, N., Hervás-Martínez, C. and Ortiz-Boyer, D. (2005) 'Cooperative coevolution of artificial neural network ensembles for pattern classification', *IEEE Transactions on evolutionary computation*, 9(3), pp. 271-302.
- Gibert, K., Sanchez-Marre, M. and Codina, V. (2010) 'Choosing the right data mining technique: classification of methods and intelligent recommendation', *International Environmental Modelling and Software Society*, .
- Gillespie, T. (1981) 'Technical considerations in the worldwide standardization of road roughness measurement', *University of Michigan Report to the World Bank*, .
- Gillespie, T.D. (1992a) *Fundamentals of vehicle dynamics*, .
- Gillespie, T.D. (1992b) *Everything you always wanted to know about the iri, but were afraid to ask*.
- González, A., O'brien, E.J., Li, Y. and Cashell, K. (2008) 'The use of vehicle acceleration measurements to estimate road roughness', *Vehicle System Dynamics*, 46(6), pp. 483-499.
- Greibe, P. (2007) 'Braking distance, friction and behaviour', *Trafitec, Scion-DTU*, .
- Haas, R., Hudson, W.R. and Zaniewski, J.P. (1994) *Modern pavement management*.
- Halcrow Group Limited (2011) 'Advice to Local Authorities: Using SCANNER survey results', *SCANNER surveys for local roads*, 3.
- Hall, G. (2015) 'Pearson's correlation coefficient', *other words*, 1(9).
- Hall, J., Smith, K.L., Titus-Glover, L., Wambold, J.C., Yager, T.J. and Rado, Z. (2009) 'Guide for pavement friction', *Final Report for NCHRP Project*, 1, pp. 43.
- Han, J., Pei, J. and Kamber, M. (2011) *Data mining: concepts and techniques*. Elsevier.

- Hand, D.J., Mannila, H. and Smyth, P. (2001) *Principles of data mining*. MIT press.
- Hassan, R. (2015) 'Modelling bituminous surfacing distress data using logistic regression', *WIT Transactions on The Built Environment*, 146, pp. 435-446.
- Heisler, H. (2002) *Advanced vehicle technology*. Elsevier.
- Hesami, R. and McManus, K.J. (2009) *Signal processing approach to road roughness analysis and measurement*. IEEE, pp. 1.
- Hruschka, H. (1993) 'Determining market response functions by neural network modeling: A comparison to econometric techniques', *European Journal of Operational Research*, 66(1), pp. 27-35.
- Ignatov, D. and Ignatov, A. (2017) *Decision Stream: Cultivating Deep Decision Trees*. IEEE, pp. 905.
- Imandoust, S.B. and Bolandraftar, M. (2013) 'Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background', *International Journal of Engineering Research and Applications*, 3(5), pp. 605-610.
- Islam, S., Buttlar, W.G., Aldunate, R.G. and Vavrik, W.R. (2014) *Use of Cellphone Application to Measure Pavement Roughness*. pp. 553.
- Jang, J., Yang, Y., Smyth, A.W., Cavalcanti, D. and Kumar, R. (2016) 'Framework of Data Acquisition and Integration for the Detection of Pavement Distress via Multiple Vehicles', *Journal of Computing in Civil Engineering*, , pp. 04016052.
- Jones, H. and Forslof, L. (2014) *Roadroid continuous road condition monitoring with smart phones*.
- Juba, B. and Le, H.S. (2017) 'Precision-Recall versus Accuracy and the Role of Large Data Sets', .
- Kaczalek, B. and Borkowski, A. (2016) 'URBAN ROAD DETECTION IN AIRBORNE LASER SCANNING POINT CLOUD USING RANDOM FOREST ALGORITHM', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B3.
- Kargah-Ostadi, N. (2014) 'Comparison of Machine Learning Techniques for Developing Performance Prediction Models' *Computing in Civil and Building Engineering (2014)*, pp. 1222-1229.
- Kashani, A.T. and Mohaymany, A.S. (2011) 'Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models', *Safety Science*, 49(10), pp. 1314-1320.

Kearns, M.J. and Valiant, L.G. (1988) *Learning Boolean formulae or finite automata is as hard as factoring*. Harvard University, Center for Research in Computing Technology, Aiken Computation Laboratory.

Kerali, H. (2002) 'Road Asset Management', *Road Asset Management Principles*, , pp. 1-13.

Kerali, H., Snaith, M.S. and Boyce, A.M. (1990) *Investment appraisal in road maintenance management*. UK: Available at:(Accessed: .

Kerali, H.G., Odoki, J. and Stannard, E. (2006) 'Overview of HDM-4', *The Highway Development and Management Series, published by World Road Association and The World Bank*, 1.

Kerali, H., Robinson, R. and Paterson, W. (1998) *Role of the new HDM-4 in highway management*. Citeseer, pp. 1.

Kionix (2015) *AN012 Accelerometer Errors*. Available at:
<http://kionixfs.kionix.com/en/document/AN012%20Accelerometer%20Errors.pdf>
(Accessed: 27/07/2018).

KNIME (2018) *KNIME open for innovation*. Available at: <https://www.knime.com/>
(Accessed: 07/08/2018).

KNIME (2015) *Seven Techniques for Data Dimensionality Reduction*. Available at:
<https://www.knime.com/blog/seven-techniques-for-data-dimensionality-reduction>
(Accessed: 07/08/2018).

Koichi, Y. (2014) *Collecting Pavement Big Data by using Smartphone*. Bali, Indonesia: IRF. Available at:(Accessed: .

Kos, A., Tomažič, S. and Umek, A. (2016a) 'Evaluation of smartphone inertial sensor performance for cross-platform mobile applications', *Sensors*, 16(4), pp. 477.

Kos, A., Tomažič, S. and Umek, A. (2016b) 'Suitability of smartphone inertial sensors for real-time biofeedback applications', *Sensors*, 16(3), pp. 301.

Kotsiantis, S.B., Zaharakis, I. and Pintelas, P. (2007) 'Supervised machine learning: A review of classification techniques', *Emerging artificial intelligence applications in computer engineering*, 160, pp. 3-24.

Kouzani, A. (2007) *Road-sign identification using ensemble learning*. Institute of Electrical and Electronics Engineers (IEEE), pp. 438.

Kropáč, O. and Múčka, P. (2005) 'Be careful when using the International Roughness Index as an indicator of road unevenness', *Journal of Sound and Vibration*, 287(4), pp. 989-1003.

Kuan, C. and White, H. (1994) 'Artificial neural networks: an econometric perspective*', *Econometric Reviews*, 13(1), pp. 1-91.

Levulytė, L., Žuraulis, V. and Sokolovskij, E. (2014) 'THE RESEARCH OF DYNAMIC CHARACTERISTICS OF A VEHICLE DRIVING OVER ROAD ROUGHNESS', *Eksploatacja i Niezawodność – Maintenance and Reliability*, 16(4), pp. 518-525.

Lewis-Beck, C. and Lewis-Beck, M.S. (2015) *Applied regression: An introduction*. Sage publications.

Liaw, A. and Wiener, M. (2002) 'Classification and regression by randomForest', *R news*, 2(3), pp. 18-22.

Liu, Y., Li, X. and Shao, X. (2007) *Short-term traffic flow prediction based on Lagrange support vector regression*. pp. 1249.

Loh, W.Y. (2011) 'Classification and regression trees', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), pp. 14-23.

Loizos, A. and Plati, C. (2008) 'An alternative approach to pavement roughness evaluation', *International Journal of Pavement Engineering*, 9(1), pp. 69-78.

Loprencipe, G. and Cantisani, G. (2013) 'Unified analysis of road pavement profiles for evaluation of surface characteristics', *Modern Applied Science*, 7(8), pp. 1.

Marcondes, J.A., Snyder, M.B. and Singh, S.P. (1992) 'Predicting vertical acceleration in vehicles through road roughness', *Journal of Transportation Engineering*, 118(1), pp. 33-49.

Markham, I.S., Mathieu, R.G. and Wray, B.A. (2000) 'Kanban setting through artificial intelligence: a comparative study of artificial neural networks and decision trees', *Integrated Manufacturing Systems*, 11(4), pp. 239-246.

Mataei, B., Zakeri, H., Zahedi, M. and Nejad, F.M. (2016) 'Pavement friction and skid resistance measurement methods: a literature review', *Open Journal of Civil Engineering*, 6(04), pp. 537.

McGhee, K.H. (2004) *Automated pavement distress collection techniques*. Transportation Research Board.

McKnight, P.E. and Najab, J. (2010) *Mann-Whitney U Test*.

McClean, J. and Ramsay, E. (1996) *Interpretations of road profile-roughness data: Review and research needs*. Australia: ARRB Transport Research Ltd. Available at: (Accessed: .

Mechanic Simulation (2017) *CarSim*. Available at: <https://www.carsim.com/company/customers/carsim.php#suppliers> (Accessed: .

- Mednis, A., Strazdins, G., Zviedris, R., Kanonirs, G. and Selavo, L. (2011) *Real time pothole detection using android smartphones with accelerometers*. IEEE, pp. 1.
- Meegoda, J.N. and Gao, S. (2014) 'Roughness Progression Model for Asphalt Pavements Using Long-Term Pavement Performance Data', *Journal of Transportation Engineering*, 140(8).
- Melhem, H.G. and Cheng, Y. (2003) 'Prediction of remaining service life of bridge decks using machine learning', *Journal of Computing in Civil Engineering*, 17(1), pp. 1-9.
- Menard, S. (2002) *Applied logistic regression analysis*. Sage.
- Mertz, C., Varadharajan, S., Jose, S., Sharma, K., Wander, L. and Wang, J. (2014) *City-wide road distress monitoring with smartphones*. pp. 1.
- Morosiuk, G., Riley, M. and Odoki, J. (2004) 'Modelling road deterioration and works effects-version 2-Highway Development and Management-HDM-4', *Highway development and management series*, .
- Naranjo, C.C.M. (2008) 'Analysis and modeling of MEMS based inertial sensors', *Kungliga Tekniska Kgskolan, Stockholm, School of Electrical Engineering*, .
- Nayak, R., Emerson, D., Weligamage, J. and Piyatrapoomi, N. (2011) *Road crash proneness prediction using data mining*. ACM, pp. 521.
- Nitsche, P., Stütz, R., Kammer, M. and Maurer, P. (2012) 'Comparison of Machine Learning Methods for Evaluating Pavement Roughness Based on Vehicle Response', *Journal of Computing in Civil Engineering*, 28(4), pp. 04014015.
- NovAtel (2014) *IMU Errors and Their Effects*. Available at: <https://www.novatel.com/assets/Documents/Bulletins/APN064.pdf> (Accessed: 27/08/2018).
- Nsofor, G.C. (2006) *A Comparative Analysis of Predictive Data-Mining Techniques*.
- Odoki, J. and Kerali, H.G. (2000) 'Analytical framework and model descriptions', *The Highway Development and Management Series*, 4.
- Olinsky, A., Kennedy, K. and Brayton Kennedy, B. (2014) 'Assessing gradient boosting in the reduction of misclassification error in the prediction of success for actuarial majors', *Case Studies In Business, Industry And Government Statistics*, 5(1), pp. 12-16.
- Organiscak, M.J. (2014) *Model Based Suspension Calibration for Hybrid Vehicle Ride and Handling Recovery*. . The Ohio State University.
- Paefgen, J., Kehr, F., Zhai, Y. and Michahelles, F. (2012) *Driving behavior analysis with smartphones: insights from a controlled field study*. ACM, pp. 36.

Papagiannakis, A. and Delwar, M. (2001) *Incorporating user costs into pavement management decisions*.

Parker, C. (2011) *An analysis of performance measures for binary classifiers*. IEEE, pp. 517.

Paterson, W.D. and Scullion, T. (1990) *Information systems for road management: draft guidelines on system design and data issues*. Washington: World Bank.

Peckov, A. (2012) *A Machine Learning Approach to Polynomial Regression*. . PhD thesis, Jozef Stefan International Postgraduate School, Ljubljana.

Pelckmans, K., Suykens, J.A., Van Gestel, T., De Brabanter, J., Lukas, L., Hamers, B., De Moor, B. and Vandewalle, J. (2002) 'LS-SVMLab: a matlab/c toolbox for least squares support vector machines', *Tutorial.KULeuven-ESAT.Leuven, Belgium*, 142, pp. 1-2.

Perlich, C., Provost, F. and Simonoff, J.S. (2003) 'Tree induction vs. logistic regression: A learning-curve analysis', *Journal of Machine Learning Research*, 4(Jun), pp. 211-255.

Pisano, P. and Goodwin, L. (2003) 'Best Practices For Road Weather Management, Version 2.0', .

Qiu, H., Chen, J., Jain, S., Jiang, Y., McCartney, M., Kar, G., Bai, F., Grimm, D.K., Gruteser, M. and Govindan, R. (2018) 'Towards Robust Vehicular Context Sensing', *IEEE Transactions on Vehicular Technology*, 67(3), pp. 1909-1922.

Rajamohan, D., Gannu, B. and Rajan, K.S. (2015) 'MAARGHA: a prototype system for road condition and surface type estimation by fusing multi-sensor data', *ISPRS International Journal of Geo-Information*, 4(3), pp. 1225-1245.

Rawlings, J.O., Pantula, S.G. and Dickey, D.A. (2001) *Applied regression analysis: a research tool*. Springer Science & Business Media.

Razi, M.A. and Athappilly, K. (2005) 'A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models', *Expert Systems with Applications*, 29(1), pp. 65-74.

Refenes, A.N., Zapranis, A. and Francis, G. (1994) 'Stock performance modeling using neural networks: a comparative study with regression models', *Neural Networks*, 7(2), pp. 375-388.

Reynolds, J., Wesson, K., Desbiez, A.L., Ochoa-Quintero, J.M. and Leimgruber, P. (2016) 'Using Remote Sensing and Random Forest to Assess the Conservation Status of Critical Cerrado Habitats in Mato Grosso do Sul, Brazil', *Land*, 5(2), pp. 12.

Ripley, B. (1996) 'Neural networks and pattern recognition', *Cambridge University*, .

Roadroid (2013) *Roadroid-Road Conditioning Monitoring Using Smart phones*. Available at:
<http://www.roadroid.se/common/References/Roadroid%20Quick%20Start%20-%20Ver%201.2.1%20Nov%202013.pdf> (Accessed: .

Robinson, R. (2008) *Restructuring road institutions, finance and management: volume 1: concepts and principles*.

Robinson, R., Danielson, U. and Snaith, M. (1998) *Road Maintenance Management-Concepts and Systems*.

Rogers, M.J., Hrovat, K., McPherson, K., Moskowitz, M.E. and Reckart, T. (1997) 'Accelerometer data analysis and presentation techniques', .

Rohrer, B. (2016) *Machine learning algorithm cheat sheet for Microsoft Azure Machine Learning Studio*. Available at:
<https://azure.microsoft.com/en-gb/documentation/articles/machine-learning-algorithm-cheat-sheet/> (Accessed: .

Rong, C. and Chun, W. (2010) *Prediction Model for Urban Expressway Short-Term Traffic Flow Based on the Support Vector Regression*. ASCE, pp. 2374.

Roy, K., Das, R.N., Ambure, P. and Aher, R.B. (2016) 'Be aware of error measures. Further studies on validation of predictive QSAR models', *Chemometrics and Intelligent Laboratory Systems*, 152, pp. 18-33.

Saaty, T.L. (1990) 'The analytic hierarchy process', *European Journal of Operational Research*, 48, pp. 9-26.

Saeedi, S. and El-Sheimy, N. (2015) 'Activity Recognition Using Fusion of Low-Cost Sensors on a Smartphone for Mobile Navigation Application', *Micromachines*, 6(8), pp. 1100-1134.

Santibanez, S.F., Kloft, M. and Lakes, T. (2014) 'Performance Analysis of Machine Learning Algorithms for Regression of Spatial Variables. A Case Study in the Real Estate Industry', .

Sayers, M.W. (1995) 'On the calculation of international roughness index from longitudinal road profile', *Transportation Research Record*, (1501).

Sayers, M.W., Gillespie, T.D. and Paterson, W.D. (1986) *Guidelines for conducting and calibrating road roughness measurements*.

Sayers, M.W. and Karamihas, S.M. (1998) 'The little book of profiling', *Ann Arbor: Transportation Research Institute, University of Michigan*. http://www.umtri.umich.edu/erd/roughness/lit_book.pdf .

- Sayers, M., Gillespie, T. and Queiroz, C. (1986) 'The international road roughness experiment: A basis for establishing a standard scale for road roughness measurements', *Transportation Research Record*, 1084, pp. 76-85.
- Schapire, R.E. (2003) 'The boosting approach to machine learning: An overview' *Nonlinear estimation and classification* Springer, pp. 149-171.
- Schlotjes, M.R. (2013) *The development of a diagnostic approach to predicting the probability of road pavement failure*. . ResearchSpace@ Auckland.
- Scholotjes, M., Visser, A. and Bennett, C. (2014) 'Evaluation of a smartphone roughness meter', .
- Seyfi, M., Rawat, R., Weligamage, J. and Nayak, R. (2013) 'A data analytics case study assessing factors affecting pavement deflection values', *Int.J.Business Intelligence and Data Mining*, 8(3), pp. 199.
- SFL Scientific (2015) *Time-Series Analysis: Wearable Devices Using Dtw And Knn*. Available at: <https://sflscientific.com/case-studies/2016/6/4/time-series-analysis-fitbit-using-dtw-and-knn> (Accessed: .
- Shahnazari, H., Tutunchian, M.A., Mashayekhi, M. and Amini, A.A. (2012) 'Application of soft computing for prediction of pavement condition index', *Journal of Transportation Engineering*, 138(12), pp. 1495-1506.
- Shrestha, D.L. and Solomatine, D.P. (2006) 'Experiments with AdaBoost. RT, an improved boosting scheme for regression', *Neural computation*, 18(7), pp. 1678-1710.
- Simmons, R. (1998) *Calculating Grms*. Available at: <https://femci.gsfc.nasa.gov/random/randomgrms.html> (Accessed: 17/08/2018).
- Singh, S. (2003) 'Stratified and Post-Stratified Sampling' *Advanced Sampling Theory with Applications* Springer, pp. 649-764.
- Snaith, M. (1998) *The development and implementation of pavement management systems: a case study over 15 years*.
- Snaith, M., Tillotson, H., Kerali, H. and Wilkins, A. (1994) *Knowledge-based systems for maintenance*. Citeseer, .
- So, J., Park, B., Wolfe, S.M. and Dedes, G. (2014) 'Development and validation of a vehicle dynamics integrated traffic simulation environment assessing surrogate safety', *Journal of Computing in Civil Engineering*, 29(5), pp. 04014080.
- Sokolova, M. and Lapalme, G. (2009) 'A systematic analysis of performance measures for classification tasks', *Information Processing & Management*, 45(4), pp. 427-437.

- Soleimani, A. and sahebi, A. (2012) 'Using neural networks to predict road roughness', *Journal of Solid and Fluid Mechanics*, 2(3), pp. 63-69.
- Strazdins, G., Mednis, A., Kanonirs, G., Zviedris, R. and Selavo, L. (2011) *Towards vehicular sensor networks with android smartphones for road surface monitoring*. pp. 2015.
- Sun, L. (2001) 'Developing spectrum-based models for international roughness index and present serviceability index', *Journal of Transportation Engineering*, 127(6), pp. 463-470.
- Taylor, P., Anand, S.S., Griffiths, N., Adamu-Fika, F., Dunoyer, A. and Popham, T. (2012) *Road type classification through data mining*. ACM, pp. 233.
- Terzi, S. (2007) 'Modeling the pavement serviceability ratio of flexible highway pavements by artificial neural networks', *Construction and Building Materials*, 21(3), pp. 590-593.
- Thomas, B. (2018) *SensorLog*. Available at: <https://itunes.apple.com/gb/app/sensorlog/id388014573?mt=8> (Accessed: 20/12/2018).
- Tsoukalas, L.H. and Uhrig, R.E. (1996) *Fuzzy and neural approaches in engineering*. John Wiley & Sons, Inc.
- UK Roads Board (2009) *SCANNER surveys FOR Local Roads-User Guide and Specification Volume 3*. London: Department of Transport. Available at:(Accessed: .
- Uys, P., Els, P.S. and Thoresson, M. (2007) 'Suspension settings for optimal ride comfort of off-road vehicles travelling on roads with different roughness and speeds', *Journal of Terramechanics*, 44(2), pp. 163-175.
- Van Baren, J. (2012) 'What is Random Vibration Testing?', *Sound and Vibration*, 46(2), pp. 9.
- Vapnik, V., Golowich, S.E. and Smola, A. (1997) 'Support vector method for function approximation, regression estimation, and signal processing', *Advances in neural information processing systems*, , pp. 281-287.
- Varunjikar, T.M., Vemulapalli, P.K. and Brennan, S.N. (2012) 'Multi-Body Vehicle Dynamics Simulation Based On Measured 3D Terrain Data', .
- Vitkiena, J., Nesterovitchb, I. and Puodžiukasc, V. (2014) 'Design of Road Longitudinal Profile, According To the Fluency and Traffic Quality of Route', .
- Von Quintus, H. and Simpson, A. (2002) *Back-calculation of layer parameters for LTPP test sections, volume II: Layered elastic analysis for flexible and rigid pavements*. Federal Highway Administration. Available at: <https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/01113/01113.pdf> (Accessed: .

Walpole, S.C., Prieto-Merino, D., Edwards, P., Cleland, J., Stevens, G. and Roberts, I. (2012) 'The weight of nations: an estimation of adult human biomass', *BMC public health*, 12(1), pp. 439.

Wang, W. and Guo, F. (2016) *RoadLab: Revamping Road Condition and Road Safety Monitoring by Crowdsourcing with Smartphone App*, .

Wang, Y. (2012) 'Ordinal logistic regression model for predicting AC overlay cracking', *Journal of Performance of Constructed Facilities*, 27(3), pp. 346-353.

Ward, C.C. and Iagnemma, K. (2009) 'Speed-independent vibration-based terrain classification for passenger vehicles', *Vehicle System Dynamics*, 47(9), pp. 1095-1113.

Washington, S.P., Karlaftis, M.G. and Mannering, F. (2010) *Statistical and econometric methods for transportation data analysis*. CRC press.

Weber, P.A. (1998) 'Towards a Canadian standard for the geometric design of speed humps', *Department of Civil and Environmental Engineering*, .

Willmott, C.J. and Matsuura, K. (2005) 'Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance', *Climate research*, 30(1), pp. 79-82.

Woolson, R.F. (2008) *Wilcoxon Signed-Rank Test*.

Workman, R., Otto, A. and Irving, A. (2016) 'The use of appropriate high-tech solutions for road network and condition analysis, with a focus on satellite imagery', .

World Bank Group (2018) *World Bank Open data - cyprus*. Available at: <https://data.worldbank.org/country/cyprus> (Accessed: 26/10/2018).

World Bank Group (2016) *Measuring Rural Access: using New Technologies*. Washington, D.C.: World Bank Group. Available at:(Accessed: .

World Bank Group (2014) *World development indicators 2014*. World Bank Publications.

World Bank Group (2008) *Economic Analysis Concepts*. Available at: <http://siteresources.worldbank.org/INTROADSHIGHWAYS/Resources/338993-1115667319236/1095944-1229373148786/04HDM-4EconomicAnalysisConcepts2008-10-22.ppt> (Accessed: 18/07/2018).

Wu, W., Chen, S. and Zheng, C. (2011) 'Traffic incident duration prediction based on support vector regression', *Proceedings of the ICCTP*, 2011, pp. 2412-2421.

XE.com (2018) *XE currency chart: USD to CYP*. Available at: <https://www.xe.com/currencycharts/?from=USD&to=CYP&view=10Y> (Accessed: 26/09/2018).

Xu, D., Mohamed, A., Yong, R. and Caporuscio, F. (1992) 'Development of a criterion for road surface roughness based on power spectral density function', *Journal of Terramechanics*, 29(4), pp. 477-486.

Yan, K., Liao, H., Yin, H. and Huang, L. (2011) 'Predicting the pavement serviceability ratio of flexible pavement with support vector machines', .

Yoon, B. and Chang, H. (2014) 'Potentialities of data-driven nonparametric regression in urban signalized traffic flow forecasting', *Journal of Transportation Engineering*, 140(7), pp. 04014027.

Yu, B., Song, X., Guan, F., Yang, Z. and Yao, B. (2016) 'k-Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition', *Journal of Transportation Engineering*, 142(6), pp. 04016018.

Zeng, H., Park, H., Fontaine, M.D., Smith, B.L. and McGhee, K.K. (2014) 'Identifying Deficient Pavement Sections by Means of an Improved Acceleration-based Metric', *Transportation Research Record: Journal of the Transportation Research Board*, .

Zeng, H. (2014) *Identifying Deficient Pavement Sections using an Improved Acceleration-based Metric*.

Zhang, Y. and Haghani, A. (2015) 'A gradient boosting method to improve travel time prediction', *Transportation Research Part C: Emerging Technologies*, 58, pp. 308-324.

Zou, T., He, Y., Zhang, N., Du, R. and Gao, X. (2015) 'Short-Time Traffic Flow Forecasting Based on the K-Nearest Neighbor Model', *traffic*, 1, pp. 36.

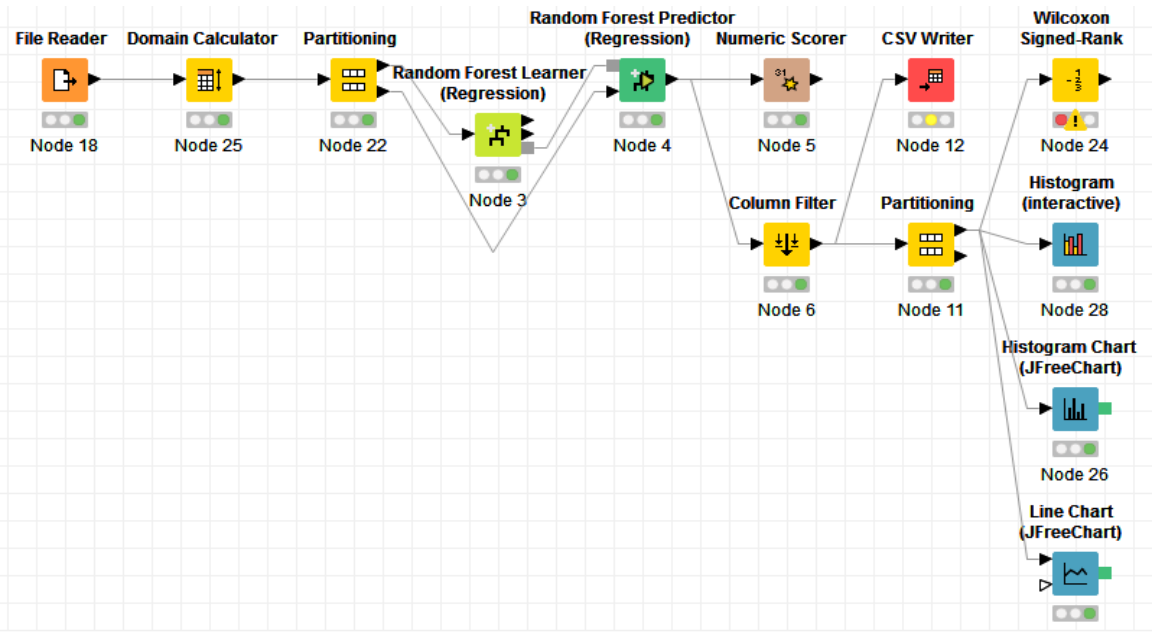
APPENDIX A

THE CONSTRUCTION OF DIFFERENT ALGORITHMS

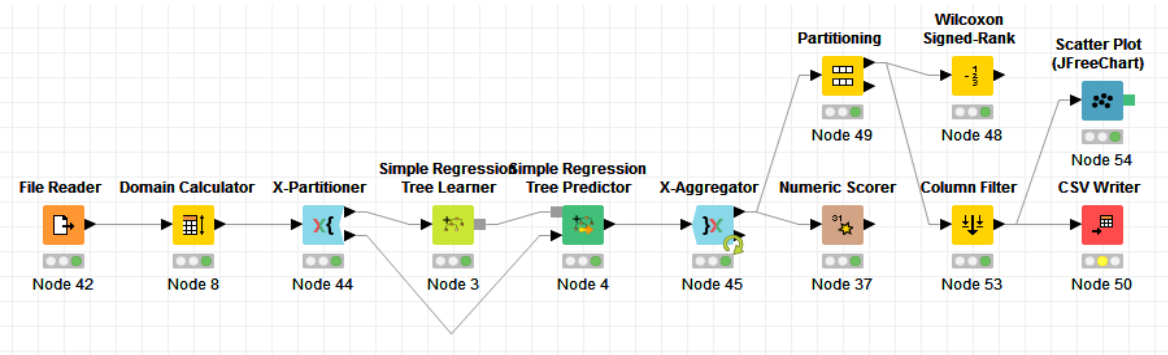
USING KNIME

Introduction

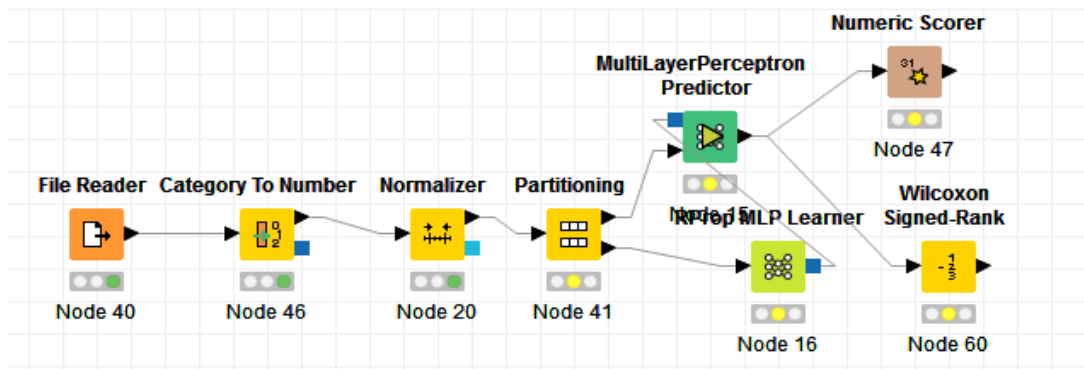
The constructions of various algorithms in KNIME platform in the research were shown as follows.



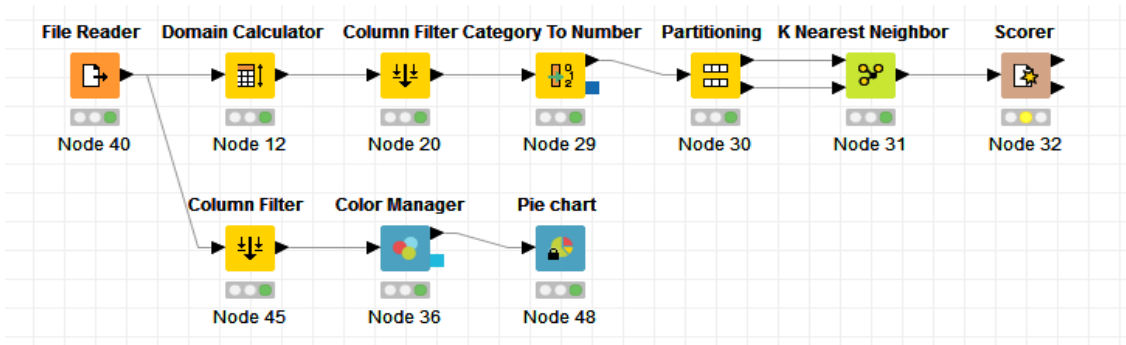
Appendix A 1: The construct of random forest algorithm in KNIME



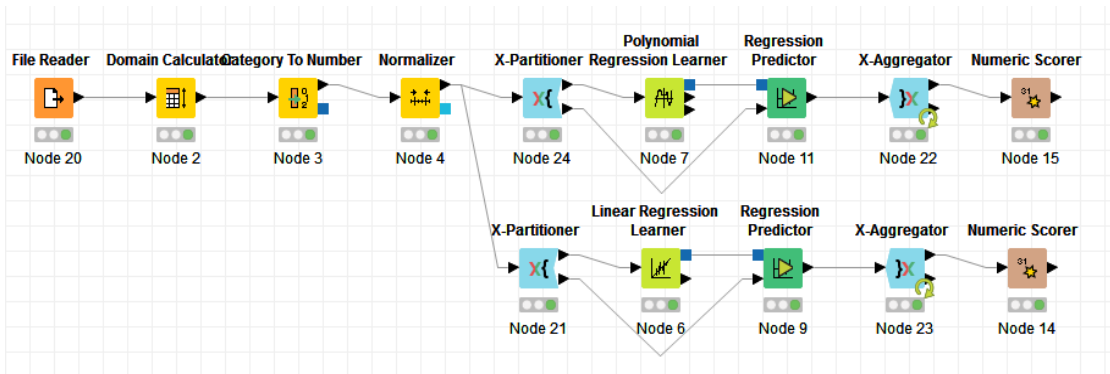
Appendix A 2: The construct of decision tree algorithm in KNIME



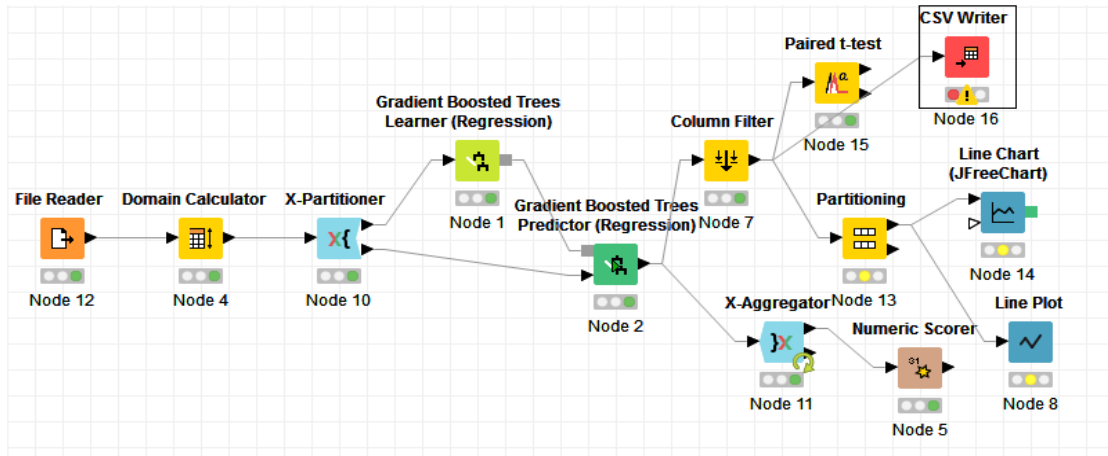
Appendix A 3: The construct of Neural Network algorithm in KNIME



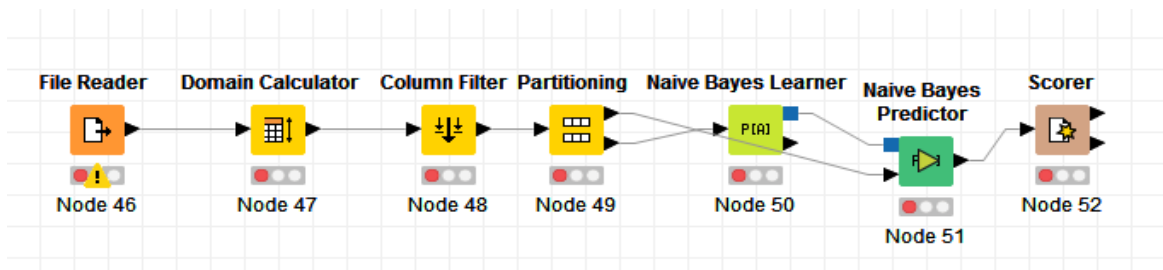
Appendix A 4: The construct of K-nearest neighbour algorithm in KNIME



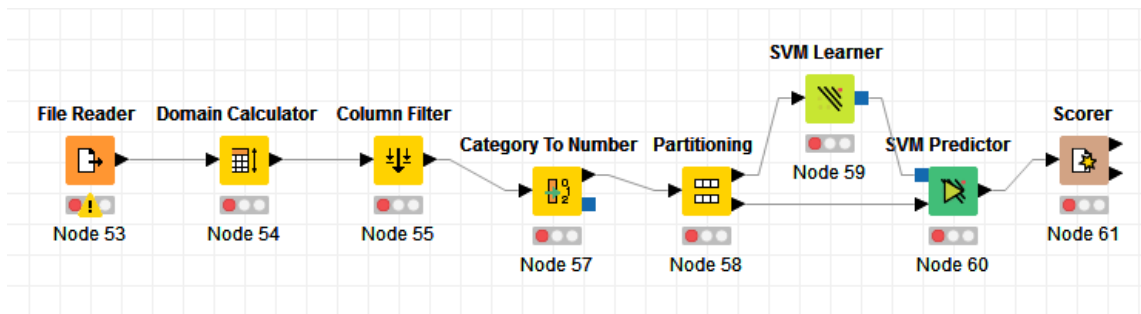
Appendix A 5: The construct of Polynomial and linear Regression in KNIME



Appendix A 6: The construct of Gradient-Boosted Tree algorithm in KNIME



Appendix A 7: The construct of Naïve Bayes algorithm in KNIME



Appendix A 8: The construct of Support vector machine algorithm in KNIME

APPENDIX B

THE USE OF SMARTPHONE DATA TO SUPPORT ROAD ASSET MANAGEMENT

INTRODUCTION

This research has shown it is possible to measure road roughness at the network level using smartphones as a platform to capture vehicle body acceleration and at potentially low cost. Indeed, **Chapter 7** developed and demonstrated smartphone-based roughness measurement systems that may potentially achieve a level of accuracy of IQL-2 (i.e. actual IRI) or IQL-3/4 (i.e. road condition classification), depending on the information available to the system. The aim of this case study is to demonstrate how road roughness information generated from the above prototype system can be used to support asset management decision-making.

Road roughness information has three main purposes in road network management (M. W. Sayers *et al.*, 1986):

- it provides a summary of the network condition on a regular basis (e.g. annually) or irregular basis (e.g. every four years),
- it can be used in models that evaluate the effectiveness of maintenance policies and pavement design standards and assess the relative road use costs, and
- roughness data can be used to identify, prioritise and optimise maintenance and rehabilitation activities on the road network.

The focus of this case study is to show explicitly how it can be used to address item 2 above. For this purpose, the road roughness data set generated in **Chapter 7** is used within HDM-4, the World Bank's standard for road investment appraisal (H. G. Kerali *et al.*, 2006). In particular, this study shows how the data can be used to set economically justifiable maintenance standards (i.e. recommended maintenance intervention levels derived from an economic analysis using HDM-4). The road network of Cyprus was used to illustrate the practical application of the approach.

Making Road Management Decisions

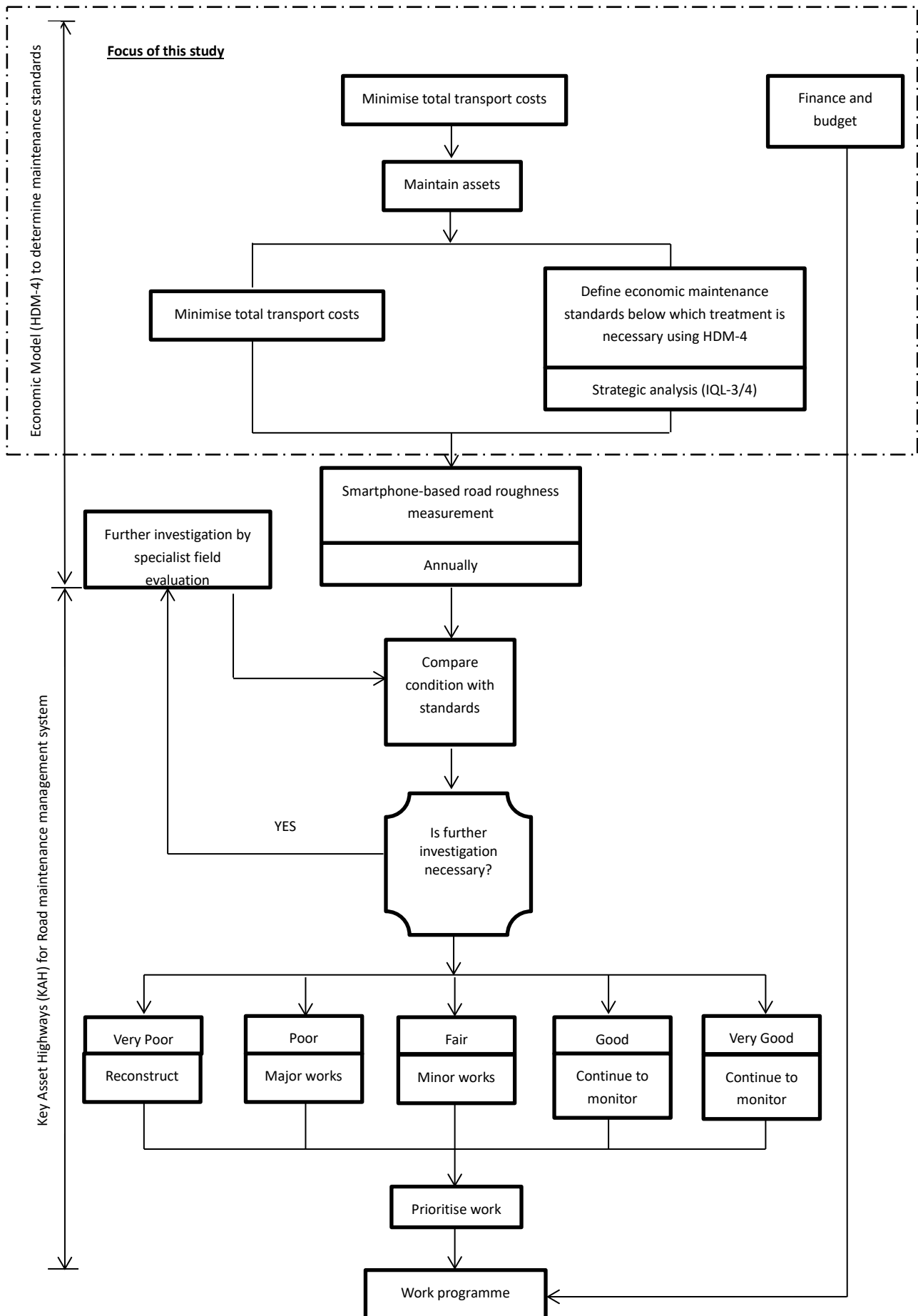
Burrow *et al.*, 2013 describe an integrated road management system which was developed for Cyprus. This system has been adapted here, as shown in **Figure 0.1**, to show how smartphone data could be used within such a system. It can be seen from the **Figure 0.1**, that the Cyprus computerised road management system operates at two levels, the network level (i.e. strategic and programming planning) and the project level (i.e. preparation and operation), incorporating economics and engineering perspectives, respectively (Snaith *et al.*, 1994). Network-level road management primarily concerns economic judgement and ranking of major road projects in which an economic model (e.g. HDM-4) can be used to set economic maintenance standards and estimate associated expenditure via a strategic analysis that ensures the entire road network performs satisfactorily (H. G. Kerali *et al.*, 2006). The road network is managed as a whole so that the maintenance standards usually represent the practice within a region, containing a set of economically justified intervention levels (H. Kerali, Snaith and Boyce, 1990). However, at the project level, maintenance engineers focus on the technical issues, including the selection of the most appropriate treatments for maintenance using a purely engineering-driven approach (Snaith, 1998). For example, a computerised road maintenance management tool (e.g. the Key

Asset Highway (KAH) in **Figure 0.1**) can be used to produce yearly reports of maintenance needs of individual sections and prioritize the order in which they should be treated. The treatment selection process is addressed by comparing the economic standards identified by the network level strategic analysis with actual road condition to identify the required remedial treatment.

Road condition data is therefore used both to set the maintenance standards and as a means by which to determine current road performance (evidently). The focus of this chapter is how road condition data obtained from a smartphone based system can be used to develop the required maintenance standards.

In many economic models, such as HDM-4, road roughness is used as the main indicator of road condition as road use costs can be directly determined from it. Typically therefore, road roughness at IQL-3/4 is used to determine the economic maintenance standards (i.e. intervention to trigger the maintenance need) via a strategic analysis, whereas at data at IQL-2 is used to monitor the current condition of individual sections. This chapter herein focuses on the former.

Figure 0.1: A systematic road maintenance management system for Cyprus (adapted from Burrow et al., 2013)



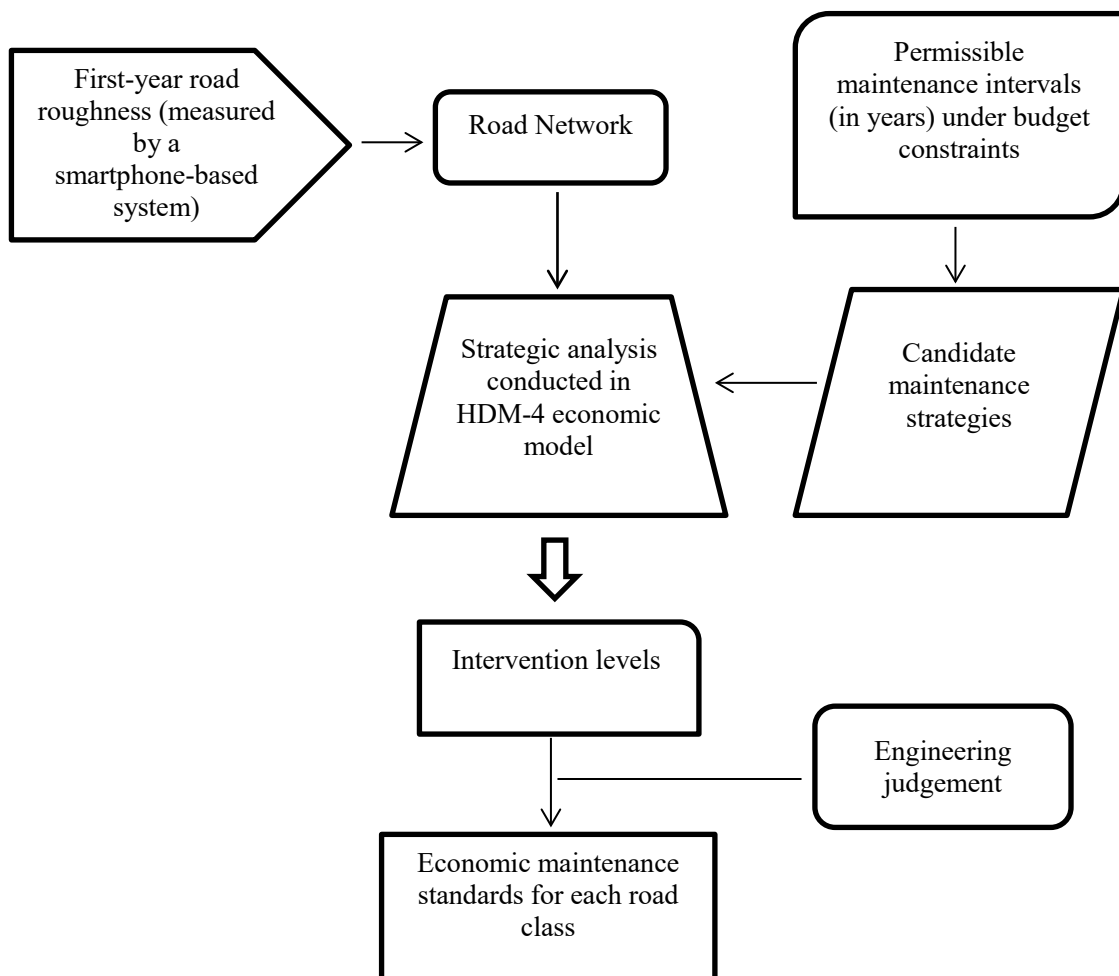
STRATEGIC ANALYSIS USING HDM-4

Many road maintenance agencies now use computerised road maintenance management systems (e.g. KAH in **Figure 0.1**) to plan their maintenance activities more efficiently and therefore achieve fair expenditure of limited resources. This, however, requires rational maintenance standards for screening defective road sections. However, the standards used have traditionally depended on the engineers' past experiences, whereas little attention has been paid to the economic returns of these standards (H. Kerali *et al.*, 1990).

An objective of carrying out a strategic analysis in HDM-4 is, therefore, to determine economic maintenance standards that minimize the total transport costs (construction, maintenance and road user costs) over a given period of time. The process is usually conducted by comparing the total transport costs associated with identified maintenance alternatives. Ideally, the maintenance alternative that provides the highest economic returns (i.e. minimum total transport costs) should be chosen.

Figure 0.2 shows the procedures for undertaking such a strategic analysis in HDM-4.

Figure 0.2: Procedure for determining economic standards using HDM-4



Cyprus road network

The road network in Cyprus (H. Kerali *et al.*, 1990) was used as an example to illustrate the procedure of utilising road roughness data to carry out a strategic analysis using HDM-4. The length of the Cyprus road network is approximately 2700 km, and all road sections are bituminous. The Cyprus road network consists of different road classes that vary from primary and tertiary roads to motorways with a wide range of conditions (Burrow *et al.*, 2013). For the purposes of the strategic analysis, the network was categorized into a number of homogeneous sub-networks. Homogeneous sub-networks were considered to be those sections of roads with similar levels of traffic, construction standards, condition and physical location. These subnetworks are considered to deteriorate similarly (Costello *et al.*, 2005). Note that within any one of these sub-networks, some variation of road conditions are likely.

To this end, the existing Cyprus road classification system was used. It consists of four classes (A, B, E and F) (i.e. sub-networks) with four conditions (R1 to R4), which categorises the entire road network into 16 combinations of representative homogeneous sections (see **Table 0.1**).

The traffic levels defined in the network matrix were based on the recorded Annual Average Daily Traffic (AADT) as reported by Burrow *et al.*, 2013 and the roads were categorized further by condition based on their IRI values measured (see **Table 0.2**), of which the representative IRI values for these homogeneous sections were assumed to be measured by the smartphone-based prototype system (i.e. at IQL-3/4) developed in **Chapter 7**, to categorise roads into very good ($0 < \text{IRI} < 2.5$), good ($2.5 < \text{IRI} < 3.5$), fair ($3.5 < \text{IRI} < 6$) or poor ($6 < \text{IRI} < 10$) condition (see **Table 8.4**).

Table 0.1: Road classification system used in the strategic analysis (H. Kerali *et al.*, 1990)

Class	Traffic volume (AADT)	Categories	Average road width (m)
A	>10,000	T1	7.5 (2 lanes)
B	5000–10,000	T2	7.0 (2 lanes)
E	1000–5000	T3	6.5 (2 lanes)
F	500–1000	T4	5.0 (2 lanes)

Roughness			
Road condition	Range (m/km)	Representative value (m/km)	Category
Very Good	0–2.5	2	R1
Good	2.5–3.5	3	R2
Fair	3.5–6	5	R3
Poor	6–10	7	R4

Table 0.2 shows the distribution matrix of road conditions for different road classes (generated based on *Figure 8.2*).

Table 0.2: Distribution matrix of road conditions for different road classes

	Class A (km)	Class B (km)	Class E (km)	Class F (km)	Total length (km)	Percentage (%)
Very Good (R1)	78	22	48	54	202	7.5
Good (R2)	186	88	183	81	538	20
Fair (R3)	243	173	707	167	1290	48
Poor (R4)	30	120	428	81	659	24.5

Total	537	403	1366	383	2689	100
--------------	------------	------------	-------------	------------	-------------	------------

Accordingly, sixteen representative road sections were constructed ranging from T1R1 to T4R4. For example, a T1R1 section has traffic level of T1 and a representative road roughness value of 2 m/km (R1; indicating good condition); while a T3R3 section is associated with traffic level of T3 and a representative road roughness value of 5 m/km (R3; indicating fair condition).

Permissible maintenance interval and maintenance strategy

The permissible maintenance interval (in years) for each homogenous road section under budget constraints is the number of years required to treat all roads on the road network with a given treatment. It is calculated by using the total budget required to apply a given treatment over all the sections divided by the allocated maintenance budget for each homogenous section. The total budget required to apply each treatment type for a road class is computed using the unit cost of that treatment type multiplied by the total area of all of the sections in the homogenous road section. The possible treatments were those used by the Cyprus Public works department and are shown together with their unit costs in **Table 0.4**. The unit costs used were based on information provided by Kerali et al., 1990, rebased to 2011.

Table 0.3: Maintenance treatments and their unit cost

Maintenance activities	Unit Cost per m ² (\$)
Surface Dressing 10 mm	4.3
Thin Overlay 20 mm	7.0
Regulating Overlay 40 mm	11.4

Thick overlay 75 mm	19.5
Edge Repair	14.4
Pothole Patching	17.2
Crack Sealing	30

The first year of analysis was assumed to be 2011. The budgets allocated for maintenance were determined using data provided by Kerali et al., 1990 factored to take into account the average inflation rate of 3.5% from 1987 to 2011 (World Bank Group, 2018) and exchange rate of 2 to US. Dollars (XE.com, 2018). **Table 0.4** shows, for example, that based on the calculated unit cost for a regulating overlay and the allocated budget of \$ 5.71 million for Class A, it would take eight years to cover the 537 km of Class A roads entirely (i.e. $537,000 \text{ m (road length)} \times 7.5 \text{ m (width)} \times \$11.4 \text{ (unit cost per m}^2) / \5.71 million).

Accordingly, **Table 0.4** shows the permissible intervention intervals for each treatment type, that the Cyprus Public Works Department uses, on each road class calculated using the allocated budget.

Table 0.4: Permissible maintenance intervals (in years) under budget constraints

TREATMENT TYPE	Class A	Class B	Class E	Class F
Surface Dressing		5	9	14
Thin Overlay	5	7	14	23
Regulating Overlay	8	12	22	38
Thick Overlay	14	21	38	64
Allocated Maintenance Budget \$ millions*	5.71	2.65	4.51	0.59

To estimate the benefit of any maintenance strategy, it is necessary to include a ‘do minimum’ alternative that enables comparison between maintenance alternatives. In the strategic analysis, routine maintenance (i.e. crack sealing, edge repair and pothole patching)

was used as the ‘do minimum’ alternative. The maintenance strategies used herein are summarised below (See **Table 0.5**).

Table 0.5: Maintenance strategies used in the strategic analysis

<i>Maintenance strategies</i>	<i>Code</i>	<i>Treatment type</i>
Do minimum (routine maintenance only)	Base alternative	Edge repairing, pothole patching and crack sealing
Alternative 1	A1	Surface dressing 10 mm + routine maintenance
Alternative 2	A2	Thin overlay 20 mm + routine maintenance
Alternative 3	A3	Regulating overlay 40 mm + routine maintenance
Alternative 4	A4	Thick overlay 75 mm + routine maintenance

HDM-4 analysis

The permissible intervention (in years) level for each treatment type for each class of road so determined were used in HDM-4 together with the road network characteristics, first-year road IRI values, vehicle and traffic information, discount rate and period of analysis to carry out the strategic analysis. As discussed above the IRI values were assumed to have been measured using smartphones collecting data to IQL-3/4.

The analysis period was set to 30 years with a discount rate of 5%. The chosen currency was U.S. dollars. The objective is to maximize Net Present Value (NPV).

The present value, known as ‘discounted value’, is the current worth of the future sums of money or cash flow based on a given specific rate of interest. Future cash flow is discounted at a particular rate. The higher the discount rate, the lower the present value of

future cash flow. Because the strategic analysis is to determine future activities, determining an appropriate discount rate is important to value the future cash flow (i.e. benefits and costs derived from road maintenance strategies). The discount rate is the rate of return for marginal public sector investments. The discount rate is normally given by the planning authority responsible for the project. The World Bank has used 10–15% for the opportunity costs of capital in developing countries (World Bank Group, 2008). Considering Cyprus as a high-income country (World Bank Group, 2014), a discount rate of 5% is used.

A number of assumptions were made before carrying out the strategic analysis including:

- Historical maintenance information (when and what treatment was applied) is assumed in accord with the measured IRI values for homogeneous sections.
- Climate, speed flow, road construction are identical for all homogeneous sections, and non-traffic information is not considered.

The results of the analysis are given below.

RESULTS

In this section, only examples of three representative sections (T1R1, T1R2 and T1R3) are investigated to demonstrate how economic maintenance standards can be set in detailed for each road class (Class A in this case).

Section: T1R1
Sensitivity: No Sensitivity Analysis Conducted

ID: T1R1
Rise + Fall: 10.00m/km

Road Class: Primary or Trunk
Width: 7.50m

Length: 78.00km
Curvature: 15.00deg/km

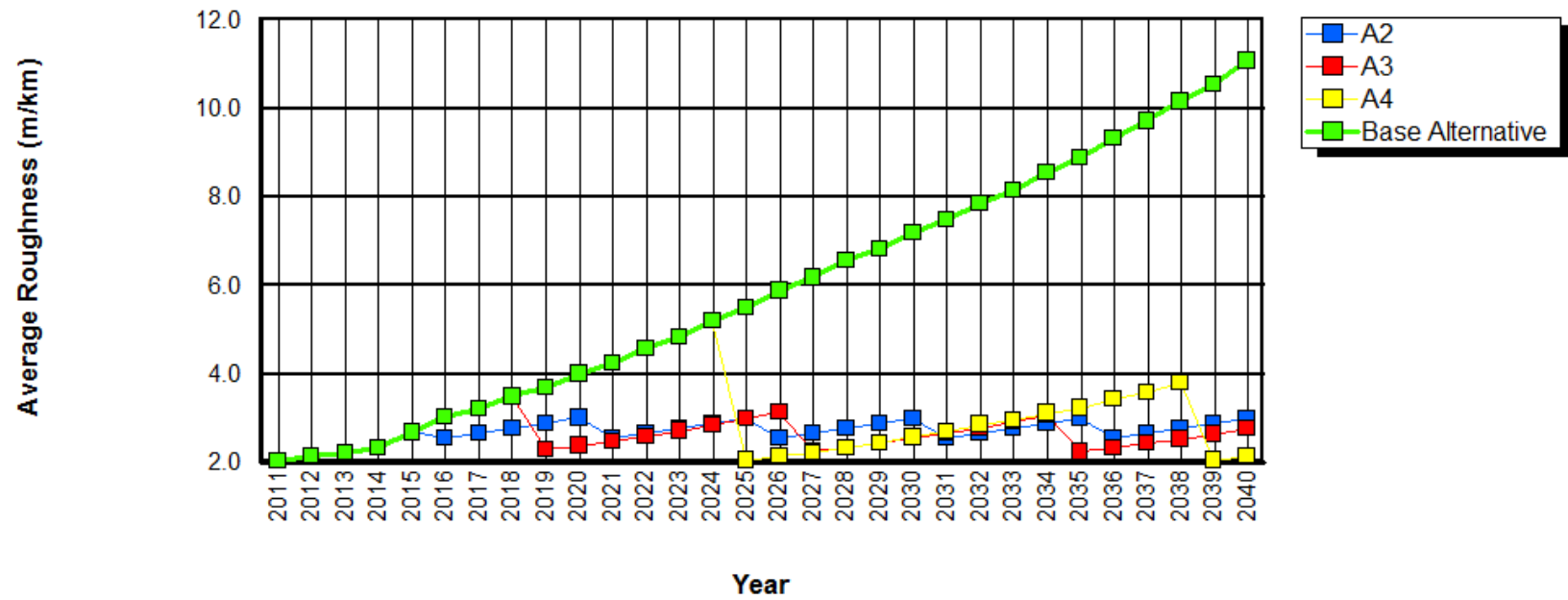


Figure 0.3: Road deterioration effect (average roughness by section) for section T1R1

Section: T1R2
Sensitivity: No Sensitivity Analysis Conducted

ID: T1R2
Rise + Fall: 10.00m/km

Road Class: Primary or Trunk
Width: 7.50m

Length: 186.00km
Curvature: 15.00deg/km

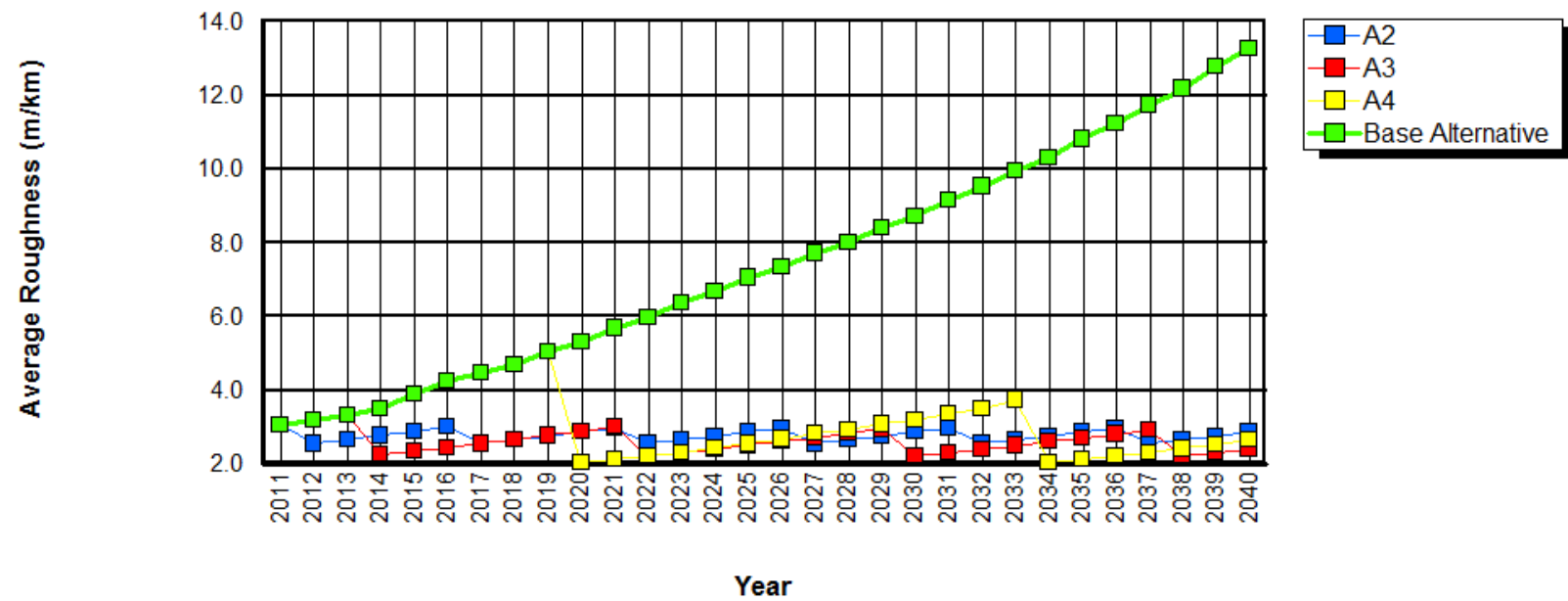


Figure 0.4: Road deterioration effect (average roughness by section) for section T1R2

HDM-4 Average Roughness by Section (Graph)

Section: T1R3
Sensitivity: No Sensitivity Analysis Conducted

ID: T1R3
Rise + Fall: 10.00m/km

Road Class: Primary or Trunk
Width: 7.50m

Length: 243.00km
Curvature: 15.00deg/km

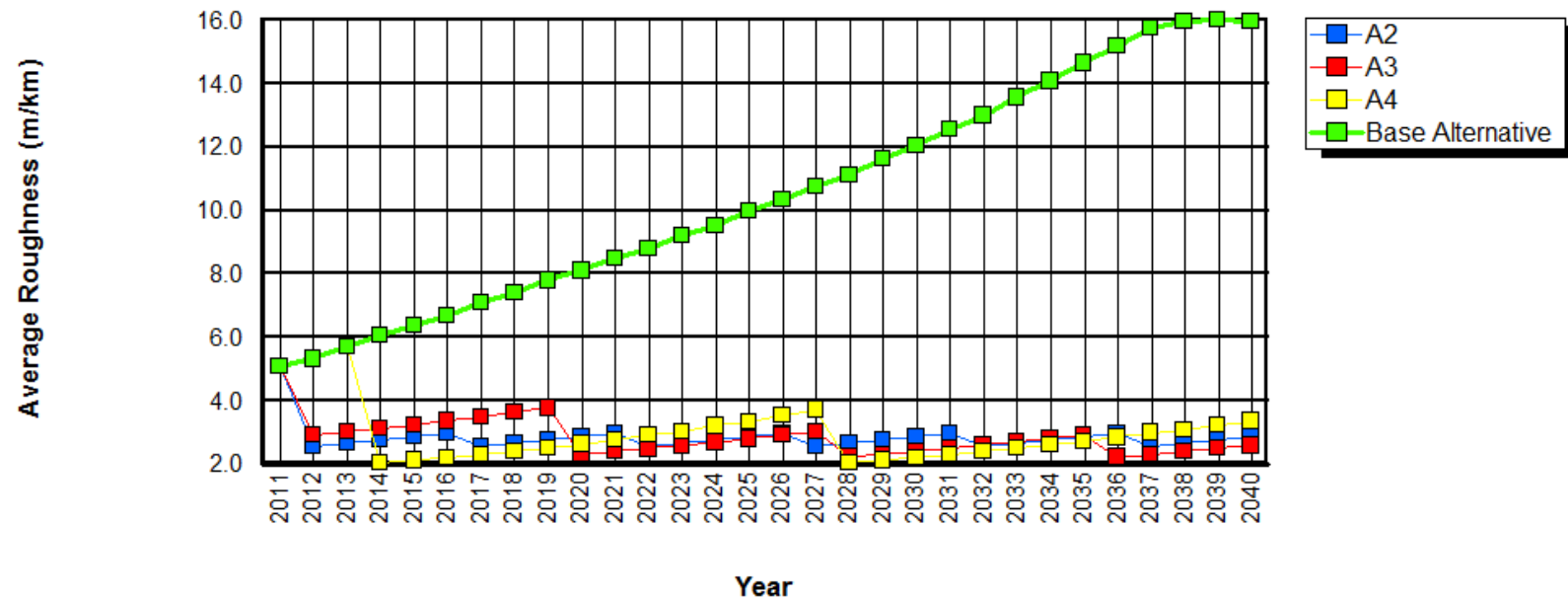


Figure 0.5: Road deterioration effect (average roughness by section) for section T1R3

Economic Indicators Summary

Study Name: C1
Run Date: 17-07-2018
Currency: US Dollar (millions)
Discount Rate: 5.00%

Section: T1R1
Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E- C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	11.811	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2	12.304	12.304	0.493	166.902	0.000	166.409	13.525	13.525	52.2 (1)
A3	9.798	9.784	-2.013	165.982	0.000	167.995	17.148	17.208	106.7 (1)
A4	17.492	11.206	5.681	144.642	0.000	138.961	7.944	12.400	104.9 (1)

Figure in brackets is number of IRR solutions in range -90 to +900

Section: T1R2
Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E- C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	33.863	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2	35.663	35.663	1.800	649.020	0.000	647.220	18.148	18.148	52.7 (1)
A3	34.031	34.031	0.168	649.977	0.000	649.809	19.095	19.095	87.3 (1)
A4	48.284	34.106	14.421	595.304	0.000	580.884	12.031	17.032	91.5 (1)

Figure in brackets is number of IRR solutions in range -90 to +900

Section: T1R3
Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E- C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	50.487	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2	46.592	46.592	-3.896	1,646.614	0.000	1,650.510	35.425	35.425	194.9 (1)
A3	49.017	49.017	-1.471	1,628.794	0.000	1,630.265	33.260	33.260	128.9 (1)
A4	72.623	59.711	22.135	1,594.964	0.000	1,572.829	21.658	26.341	99.8 (1)

Figure in brackets is number of IRR solutions in range -90 to +900

Figure 0.6: Economic Indicators Summary for sections T1R1, T1R2 and T1R3

Error! Reference source not found.Figure 0.3 to Figure 0.5 show the predicted road roughness progression for the three sections and the maintenance strategies considered. As **Figure 0.3** shows, if only the base alternative is applied (do minimum), based on current traffic, the average roughness of section T1R1 (i.e. initial IRI = 2.0 m/km) will increase to 4 m/km after 10 years (in 2021) and will continue to increase to 11 m/km (i.e. very poor condition) after 30 years (in 2041). Compared with the base alternative (do minimum), if a road is maintained periodically, the condition of section T1R1 is controlled to a level at which a treatment should perform based on predefined maintenance intervals (see **Figure 0.3**). Once a remedy treatment is applied, the condition of the representative section returns to its original condition, or close to it, after which the representative section deteriorates until the next treatment is applied. The roughness level which the representative section reaches when the treatment is applied is equal to the treatment's trigger level. For example, IRI = 3 m/km for red A3 (Regulating overlay) in **Figure 0.3**.

HDM-4 provides the Net Present Value (NPV), the Internal Rate of Return (IRR), NPV/cost ratio and total transport costs as economic measures to appraise road investment, they are demonstrated below.

Net Present Value. This value is the difference between discounted benefits and costs of a project (in this case, maintenance strategy) in which the benefits are mainly derived from savings in road user costs and sometimes also in road maintenance costs. As such, the benefit derived from a maintenance strategy is calculated by subtracting the savings in

total transport cost (i.e. the sum of construction costs, future road maintenance costs and road user costs) of such a strategy from that of a 'do minimum' strategy.

When NPV, for a given discount rate, is positive, then the alternative is accepted and vice versa. If NPV is zero, then the alternative is indifferent to the without alternative (do minimum). The higher the NPV, the greater the benefits from the investment option (maintenance alternative) compared with the base option (do minimum).

Internal Rate of Return (IRR): The NPV relies on the discount rate used in the computation of present values. A higher discount rate results in a lower NPV. The IRR of an option is the discount rate in which the present value of costs equals the present value of benefits (i.e. $NPV = 0$). It is an indicator of profitability. The higher the IRR, the better. In general, if the calculated IRR is greater than the selected discount rate, the option is economically justified.

Benefit–Cost Ratio: This is the simplest measure of profitability of an option (i.e. the amount of benefits derived for every dollar spent). It is a dimensionless index computed by dividing the calculated benefits by the discounted capital costs of an investment option (i.e. maintenance strategy) and thus represents the NPV per investment unit. It can eliminate the bias of NPV for large investment options. It is considered a measure of the efficiency of investment and can be used to rank and compare the investment options.

Total transport costs: This is the sum of construction costs, future road maintenance costs

and road user costs. Comparing the total transport costs of an option against the base option (do minimum) gives an indication of economic appraisal. Ideally, the option with the minimum total transport cost should be chosen. This is, however, the same meaning as maximised NPVs.

Figure 0.6 and **Table 0.6** summarises the economic indicators associated with the maintenance strategies that are considered for each representative section in Class A from which it may be seen that A3 is the most economically justified maintenance strategy for section T1R1, as evidenced by the highest NPV, NPV/cost ratio, IRR and total transport cost compared with the base alternative. Applying maintenance strategy A3 also showed the highest economic benefits in terms of NPV, NPV/cost ratio and total transport cost for section T1R2 compared with the other alternatives. The most economic maintenance strategy for section T1R3 is A2, as reflected by the highest NPV, NPV/cost ratio, total transport cost and IRR compared with all other alternatives.

Table 0.6: Total transport costs on Class A road representative road sections

Maintenance strategies	Total transport costs (unit: \$ millions)					
	T1R1	Difference (in %) compared with base alternative	T1R2	Difference (in %) compared with base alternative	T1R3	Difference (in %) compared with base alternative
Base alternative	1880	100	4744	100	7021	100
A2	1714	91.1	4097	86.4	5370	76.5
A3	1712	91.0	4094	86.3	5390	76.7

A4	1741	92.6	4162	87.7	5448	77.6
----	------	------	------	------	------	------

The results indicate that maintenance strategies A2 and A3 have similar economic benefits for Class A roads, as reflected by their similar Net Present Values calculated (see **Figure 0.6**). In this case, a possible maintenance strategy which would most benefit the road agency would be to choose the option associated with relatively lower road agency costs. **Figure 0.6** shows that maintenance strategy A3 generally requires less summed expenditure for maintaining all sections compared with maintenance strategy A2. Therefore, considering all the findings, maintenance strategy A3 should be selected as the economic maintenance standard for Class A roads. Applying maintenance strategy A3 decreases the total road user costs by \$166 million (discounted to present value in 2011) over 30 years for section T1R1, a 9% reduction compared with the total road user cost (discounted) estimated for the base alternative (i.e. \$1868.6 millions).

Table 0.7 shows the recommended maintenance intervention levels resulting from the HDM-4 strategic analysis.

Table 0.7: Recommended maintenance intervention levels for all road sections

Recommended maintenance activities	Roughness Intervention level (m/km)		Major deterioration intervention level (cracking in % area)
Class A roads			
1) Regulating overlay	3	AND	10
2) Routine maintenance			
Class B roads			
1) Thin overlay	3.4	AND	6
2) Routine maintenance			

Class E roads			
1) Thin overlay	4.2	AND	15
2) Routine maintenance			
Class F roads			
1) Thin overlay	7	AND	48
2) Routine maintenance			

These results are given in terms of roughness levels that would be obtained within the maintenance intervention level (years) described in **Table 0.7**. The intervention levels for other types of deterioration can only be inferred from the corresponding roughness level. For example, the critical roughness level of 3 m/km given in **Table 0.7** for Class A roads would be obtained after eight years, by which time 10% area of the road is predicted by HDM-4 to have wide cracks.

Simultaneously, following the same approach, **Table 0.7** shows the recommended intervention levels for Class B, E and F roads. The most economic maintenance standard for these road classes is a thin overlay rather than a regulating overlay.

In summary, the economic periodic maintenance standards, for example, for Class A roads, can be interpreted thus: if the roughness level reaches 3 m/km and 10% of the surface area is cracked, then regulating overlay (40 mm) should be carried out. Note that routine maintenance, as a separate item, is also included in the economic maintenance standards.

In addition, the HDM-4 also calculates the annual average road user costs per vehicle-km which provide a stranding point for road users to how much costs are spent per vehicle per km while driving on a particular route. This would therefore enable users to recognise the

benefits or loss economically derived from improved/deteriorated road condition respectively.

The results of other (apart from Class A) representative road sections of strategic analysis in HDM-4 are shown below, they will not be further analysed and demonstrated repeatedly in this section.

Section: T2R1
Sensitivity: No Sensitivity Analysis Conducted

ID: T2R1
Rise + Fall: 10.00m/km

Road Class: Primary or Trunk
Width: 7.00m

Length: 22.00km
Curvature: 15.00deg/km

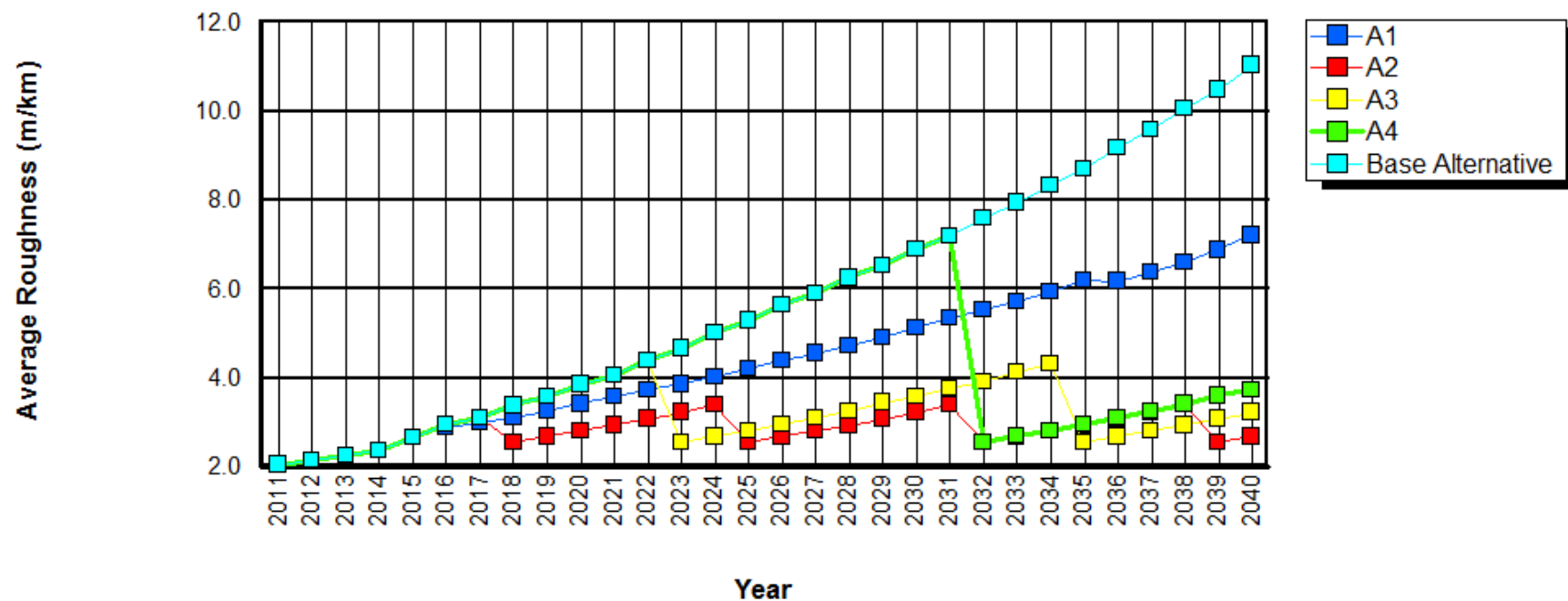


Figure 0.7: Road deterioration prediction of section T2R1

Section: T2R2
Sensitivity: No Sensitivity Analysis Conducted

ID: T2R2
Rise + Fall: 10.00m/km

Road Class: Primary or Trunk
Width: 7.00m

Length: 88.00km
Curvature: 15.00deg/km

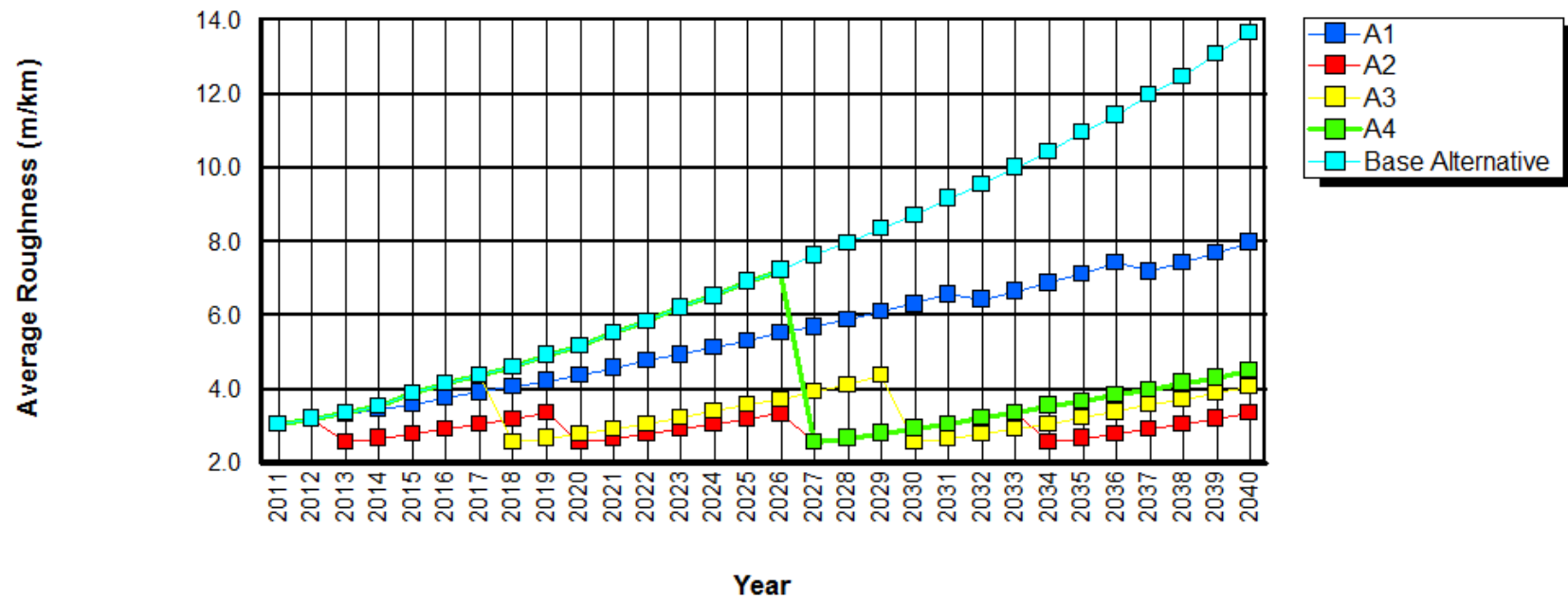


Figure 0.8: Road deterioration prediction of section T2R2

Section: T2R3
Sensitivity: No Sensitivity Analysis Conducted

ID: T2R3
Rise + Fall: 10.00m/km

Road Class: Primary or Trunk
Width: 7.00m

Length: 173.00km
Curvature: 15.00deg/km

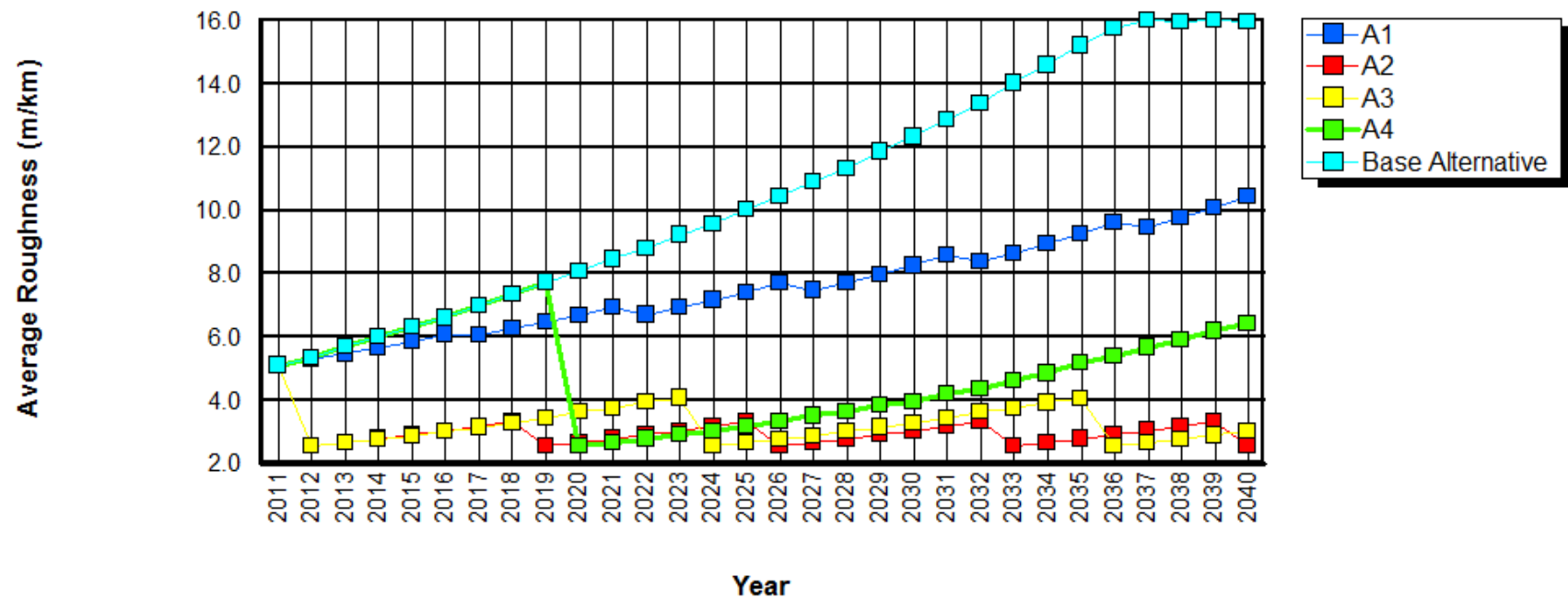


Figure 0.9: Road deterioration prediction of section T2R3

Section: T3R2
Sensitivity: No Sensitivity Analysis Conducted

ID: T3R2
Rise + Fall: 10.00m/km

Road Class: Secondary or Main
Width: 6.50m

Length: 183.00km
Curvature: 15.00deg/km

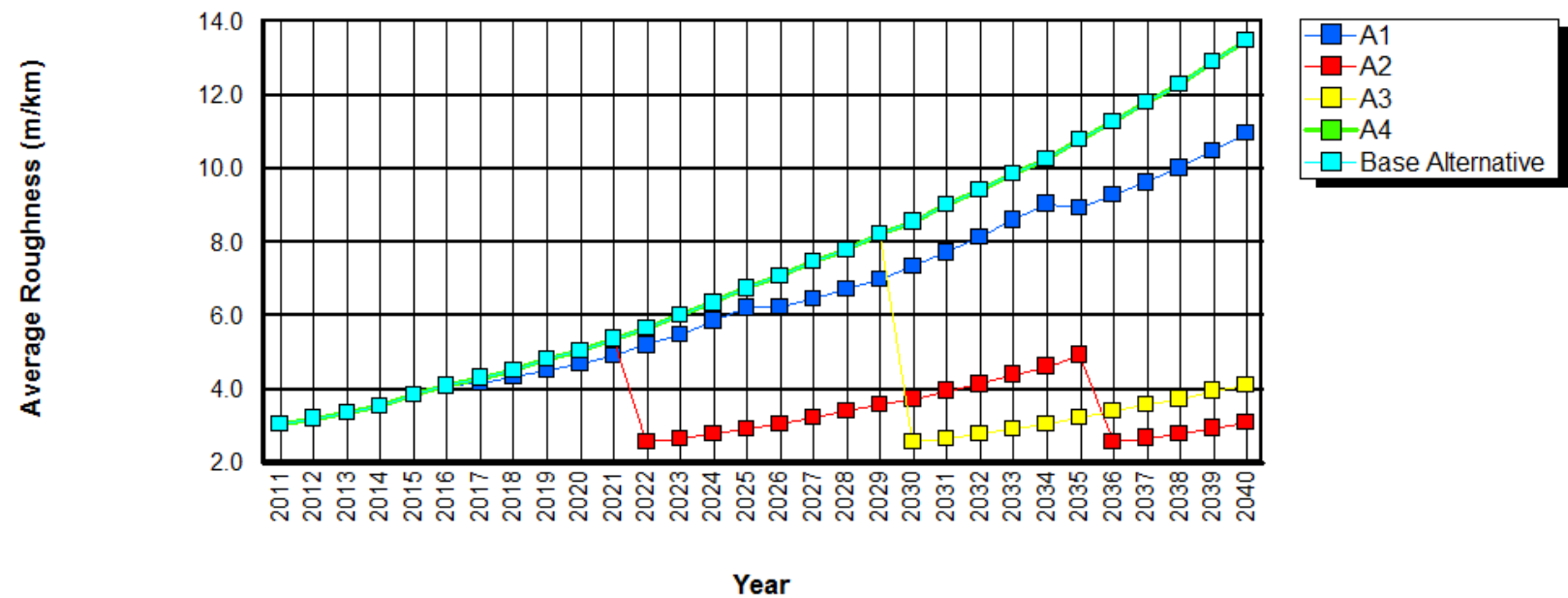


Figure 0.10: Road deterioration prediction of section T3R2

Section: T2R1

Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E - C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	3.191	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A1	1.920	1.799	-1.271	26.012	0.000	27.282	14.208	15.162	57.5 (1)
A2	2.216	2.130	-0.975	47.263	0.000	46.238	21.767	22.644	98.7 (1)
A3	3.005	1.542	-0.186	42.392	0.000	42.578	14.167	27.613	211.5 (1)
A4	3.834	1.393	0.643	29.318	0.000	28.674	7.478	20.585	212.3 (1)

Figure in brackets is number of IRR solutions in range -90 to +900

Section: T2R2

Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E - C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	15.310	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A1	9.257	8.749	-6.054	182.227	0.000	188.281	20.340	21.521	47.6 (1)
A2	12.205	11.953	-3.105	331.648	0.000	334.753	27.426	28.006	75.3 (1)
A3	13.147	7.872	-2.164	303.233	0.000	305.398	23.230	38.796	190.6 (1)
A4	18.915	7.111	3.805	240.117	0.000	236.513	12.504	33.258	193.8 (1)

Figure in brackets is number of IRR solutions in range -90 to +900

Section: T3R2
Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E- C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	29.687	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A1	15.885	7.680	-13.802	55.542	0.000	69.344	4.365	9.029	No Solution
A2	23.228	7.913	-8.459	171.322	0.000	177.781	7.654	22.466	237.0 (1)
A3	28.122	5.437	-1.565	130.142	0.000	131.707	4.683	24.225	272.9 (1)
A4	29.687	0.000	0.000	0.000	0.000	0.000	0.000	zero cost	No Solution

Figure in brackets is number of IRR solutions in range -90 to +900

H D M - 4 Economic Indicators Summary

Section: T3R3
Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E- C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	124.566	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A1	59.824	32.713	-84.742	177.655	0.000	242.396	4.052	7.410	748.5 (1)
A2	89.844	33.708	-34.722	943.425	0.000	978.147	10.887	29.020	191.8 (1)
A3	114.269	23.158	-10.297	674.523	0.000	684.820	5.993	29.572	229.2 (1)
A4	124.566	0.000	0.000	0.000	0.000	0.000	0.000	zero cost	No Solution

Figure in brackets is number of IRR solutions in range -90 to +900

Figure 0.11: Economic indicator summary of representative road sections

DISCUSSION

This case study has described how road roughness, measured by smartphone-based applications, can be used via a strategic analysis conducted in HDM-4 to support asset management decisions, particularly in determining economic maintenance standards. As demonstrated in **Chapter 7** a smartphone based system could be developed with an error of prediction of $IRI \pm 0.3$ m/km. To better understand the implications of this error a sensitivity analysis was carried out in HDM-4 to determine how the error may affect setting economic maintenance standards.

Resulting from changing current road roughness, the most important inputting data into HDM-4 that would change are those of the distribution matrix of road conditions within a network. Two scenarios were chosen to illustrate the effect of changing the distribution of road network condition of the on the determination of maintenance standards. Accordingly two new distributions of road condition were determined by adding and subtracting 0.3 m/km from the IRI values in the dataset (cf. [Section 8.2.3](#)). The resulting distributions are shown in

Table 0.8 and Table 0.9.

Table 0.8: Distribution of road condition (after adding 0.3 m/km to all IRI values in the dataset)

	Class A (km)	Class B (km)	Class E (km)	Class F (km)	Total length (km)	Percentage (%)
Very Good (R1)	38	28	96	27	188	7
Good (R2)	64	48	164	46	323	12
Fair (R3)	295	222	751	211	1479	55
Poor (R4)	140	105	355	100	699	26
Total	537	403	1366	383	2689	100

Table 0.9: Distribution of road conditions (after subtracting 0.3 m/km to all IRI values in the datasets)

	Class A (km)	Class B (km)	Class E (km)	Class F (km)	Total length (km)	Percentage (%)
Very Good (R1)	84	49	64	72	269	10
Good (R2)	212	100	249	111	672	25
Fair (R3)	215	151	639	124	1129	42
Poor (R4)	26	103	413	76	618	23
Total	537	403	1366	383	2689	100

The new distributions of road conditions were used again in an HDM-4 strategic analysis to compare the economic maintenance standards with the findings in Section 8.3. The results from these two scenarios are shown in **Figure 0.12** and **Figure 0.13**. For comparison purposes, only the results of Class A roads are presented. As can be seen in

Figure 0.12, adding an 0.3 m/km to the measured road roughness conditions within the network, the Net Present Value derived from different maintenance strategies for representative sections in ‘very good (R1)’ and ‘good (R2)’ condition decreased compared with the results shown in **Figure 0.6**, whereas the Net Present Value estimated for representative sections in ‘fair (R3)’ increased because more representative sections were predicted to have fair condition.

The above notwithstanding, after analysing the economic indicators it can be seen that, in this scenario, the economic maintenance standards are unchanged. For example, A3 was still found to be the economically justified maintenance strategy for Class A roads and the recommended intervention levels predicted for A3 also remained unchanged (See **Table 0.7**).

Further, by inspecting **Figure 0.13**, subtracting 0.3 m/km from all the IRI values of roads within the network, the Net Present Value estimated from different maintenance strategies for sections in ‘very good (R1)’ and ‘good (R2)’ condition increased compared with results in **Figure 0.6**, whereas the Net Present Value estimated for representative section in ‘fair (R3)’ condition decreased. The above notwithstanding, after carrying out an economic analysis, it was found that the economic maintenance standards predicted in this scenario remain unchanged (See **Table 0.7**). The above would therefore suggest that the roughness prediction error of the smartphone-based application does not appear to affect the determination by means of a strategic analysis, the economic maintenance standards of the

Cyprus road network.

Economic Indicators Summary

Study Name: C1

Run Date: 31-07-2018

Currency: US Dollar (millions)

Discount Rate: 5.00%

Section: T1R1

Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E - C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	5.754	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2	5.994	5.994	0.240	81.311	0.000	81.071	13.525	13.525	52.2 (1)
A3	4.773	4.757	-0.981	80.883	0.000	81.844	17.146	17.206	106.7 (1)
A4	8.522	5.459	2.768	70.467	0.000	67.699	7.944	12.400	104.9 (1)

Figure in brackets is number of IRR solutions in range -90 to +900

Section: T1R2

Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E - C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	11.652	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2	12.271	12.271	0.619	223.319	0.000	222.700	18.148	18.148	52.7 (1)
A3	11.709	11.709	0.058	223.648	0.000	223.590	19.095	19.095	67.3 (1)
A4	16.614	11.735	4.962	204.836	0.000	199.874	12.031	17.032	91.5 (1)

Figure in brackets is number of IRR solutions in range -90 to +900

Section: T1R3
Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E- C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	61.291	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2	56.562	56.562	-4.729	1,996.976	0.000	2,003.705	35.425	35.425	194.9 (1)
A3	59.506	59.506	-1.785	1,977.343	0.000	1,979.128	33.260	33.260	128.9 (1)
A4	88.163	72.489	26.872	1,936.274	0.000	1,909.401	21.658	26.341	99.8 (1)

Figure in brackets is number of IRR solutions in range -90 to +900

Figure 0.12: Economic Indicator summary for section T1R1, T1R2 and T1R3 (according to distribution matrix of road condition in

Table 0.8)

Economic Indicators Summary

Study Name: C1

Run Date: 05-08-2018

Currency: US Dollar (millions)

Discount Rate: 5.00%

Section: T1R1

Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E - C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	12.720	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2	13.250	13.250	0.530	179.741	0.000	179.210	13.525	13.525	52.2 (1)
A3	10.551	10.515	-2.168	178.750	0.000	180.919	17.148	17.208	106.7 (1)
A4	18.837	12.088	6.118	155.788	0.000	149.651	7.944	12.400	104.9 (1)

Figure in brackets is number of IRR solutions in range -90 to +900

Section: T1R2

Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E - C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	38.597	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2	40.648	40.648	2.051	739.743	0.000	737.692	18.148	18.148	52.7 (1)
A3	38.788	38.788	0.191	740.834	0.000	740.642	19.095	19.095	67.3 (1)
A4	55.033	38.873	16.436	678.519	0.000	662.082	12.031	17.032	91.5 (1)

Figure in brackets is number of IRR solutions in range -90 to +900

Section: T1R3
Sensitivity: No Sensitivity Analysis Conducted

Alternative	Present Value of Total Agency Costs (RAC)	Present Value of Agency Capital Costs (CAP)	Increase in Agency Costs (C)	Decrease in User Costs (B)	Net Exogenous Benefits (E)	Net Present Value (NPV = B + E- C)	NPV/Cost Ratio (NPV/RAC)	NPV/Cost Ratio (NPV/CAP)	Internal Rate of Return (IRR)
Base Alternative	44.870	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2	41.223	41.223	-3.447	1,456.881	0.000	1,460.328	35.425	35.425	194.9 (1)
A3	43.369	43.369	-1.301	1,441.114	0.000	1,442.416	33.260	33.260	128.9 (1)
A4	64.255	52.831	19.585	1,411.182	0.000	1,391.598	21.658	26.341	99.8 (1)

Figure in brackets is number of IRR solutions in range -90 to +900

Figure 0.13: Economic Indicator summary for section T1R1, T1R2 and T1R3 (according to distribution matrix of road condition in Table 0.9)

SUMMARY

The developed prototype system can be explored to support road asset management decisions by:

- Enabling the monitoring the road condition of each road section on a regular basis
- Facilitating the determination of economic maintenance standards
- Enabling the road user costs and road agency costs to be estimated for the derived maintenance standards
- Helping to identify the maintenance needs for monitored road sections

Error of measuring the data using smartphone prototype system affects the Information quality (at IQL-3/4) as a result the economic benefit estimated can change however this does not affect the determination of economic maintenance standards

APPENDIX C

Appendix C demonstrated the detailed analysis carried out for each sections of the measurement described in the **Chapter 8**.

Section C-1: Comparative analysis of all subsections

In order to provide additional insight into the 12,840 m road section measured, the entire road is divided into 6 sections with various of lengths from 1700 m to 3000 m.

Section One

The first section is 2300 m in length and start, end points were shown in **Figures C.1** to **C.3** below.



Figure C.1: The first part of road in Section One (start, end points)



Figure C.2: The second part of road in Section One (starting, ending points)



Figure C.3: The third part of road in Section One (starting, ending points)

The measured IRI from Roadroid and SCANNER were compared in the both time domain and in the frequency domain, as shown in **Figures C.4 to C.6**. **Table C.1** gives a summary of statistical analysis.

It can be seen from **Table C.1** that **Run 1** generally appears to correlate relatively closer to the SCANNER data, with an average IRI of 2.12 m/km that higher than SCANNER data. However, by inspecting the **Figure C.4** to **C.6** it was found that the Roadroid and SCANNER data does not agree well. However, three runs of Roadroid tests do perform a level of repeatability.

Table C.1: The statistical analysis of IRI measurements from Roadroid and SCANNER

	SCANNER	Roadroid Application		
		Run 1	Run 2	Run 3
<i>Mean Estimated IRI (m/km)</i>	<i>1.864</i>	<i>2.12</i>	<i>3.05</i>	<i>3.26</i>
<i>Variance (m/km)</i>	<i>-</i>	<i>0.25</i>	<i>1.19</i>	<i>1.39</i>
<i>Mean Ratio</i>	<i>1</i>	<i>1.58</i>	<i>2.24</i>	<i>2.44</i>

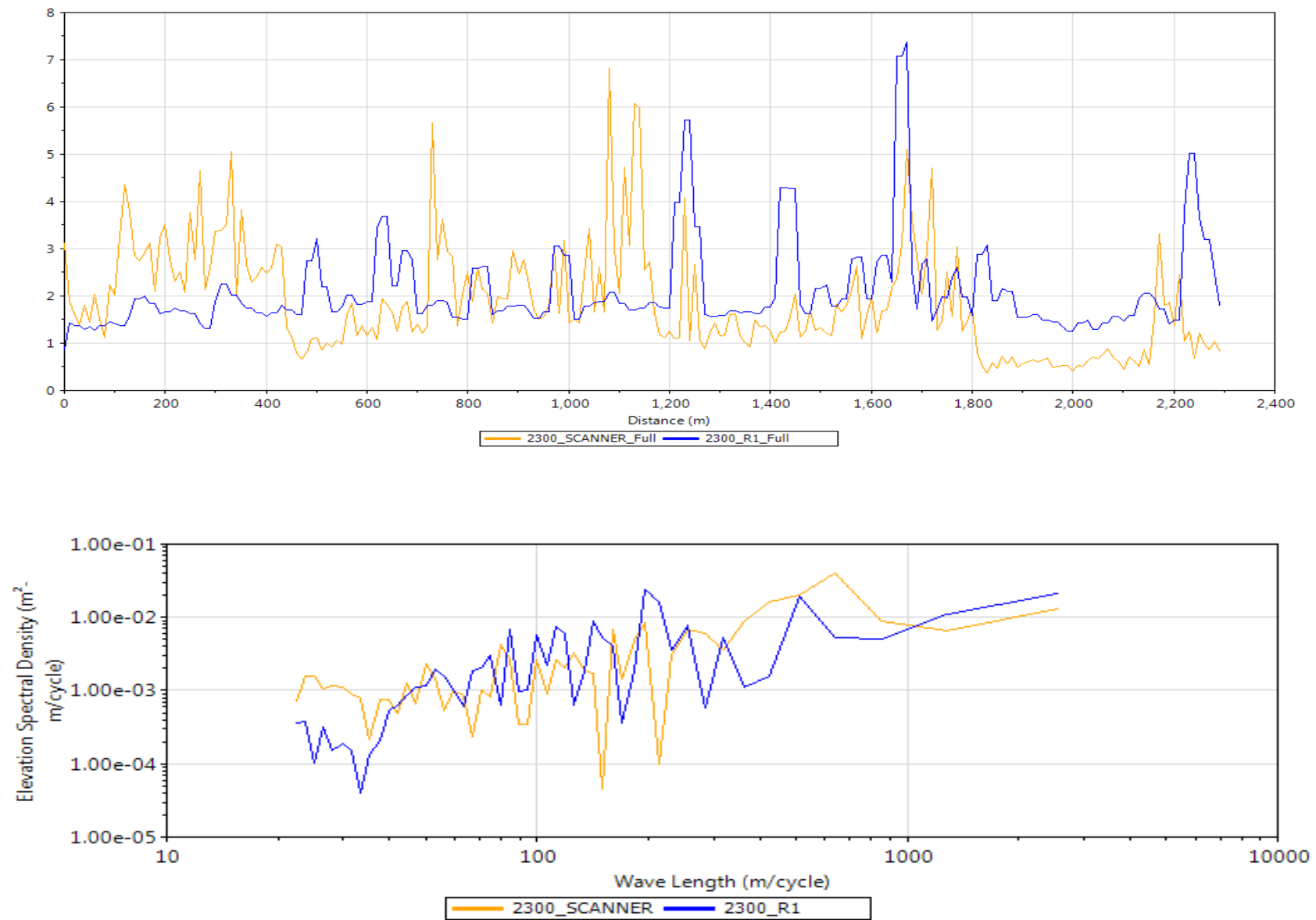


Figure C.4: Spectral Analysis of Roadroid Run 1 and SCANNER

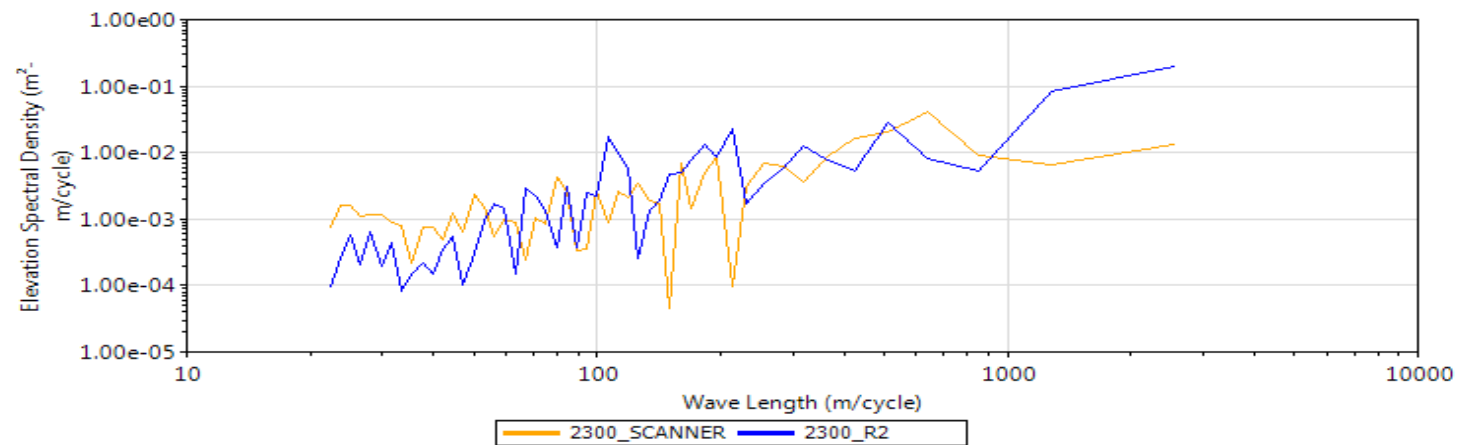
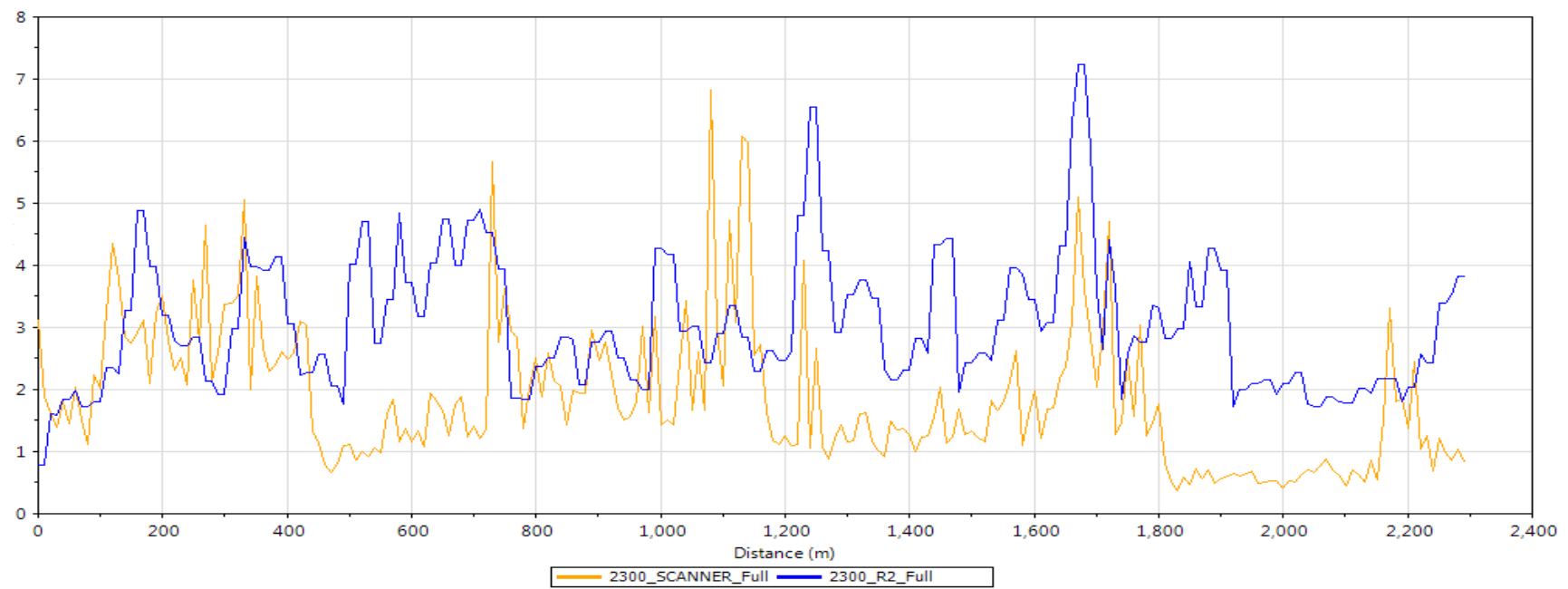


Figure C.5: Spectral Analysis of Roadroid Run 2 and SCANNER

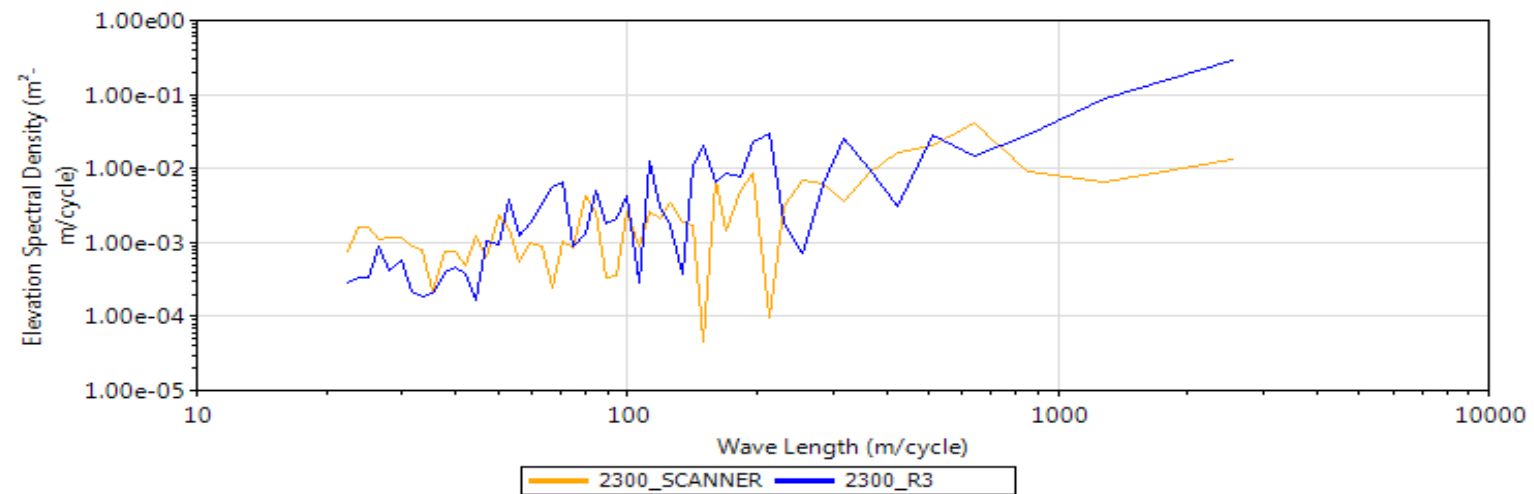
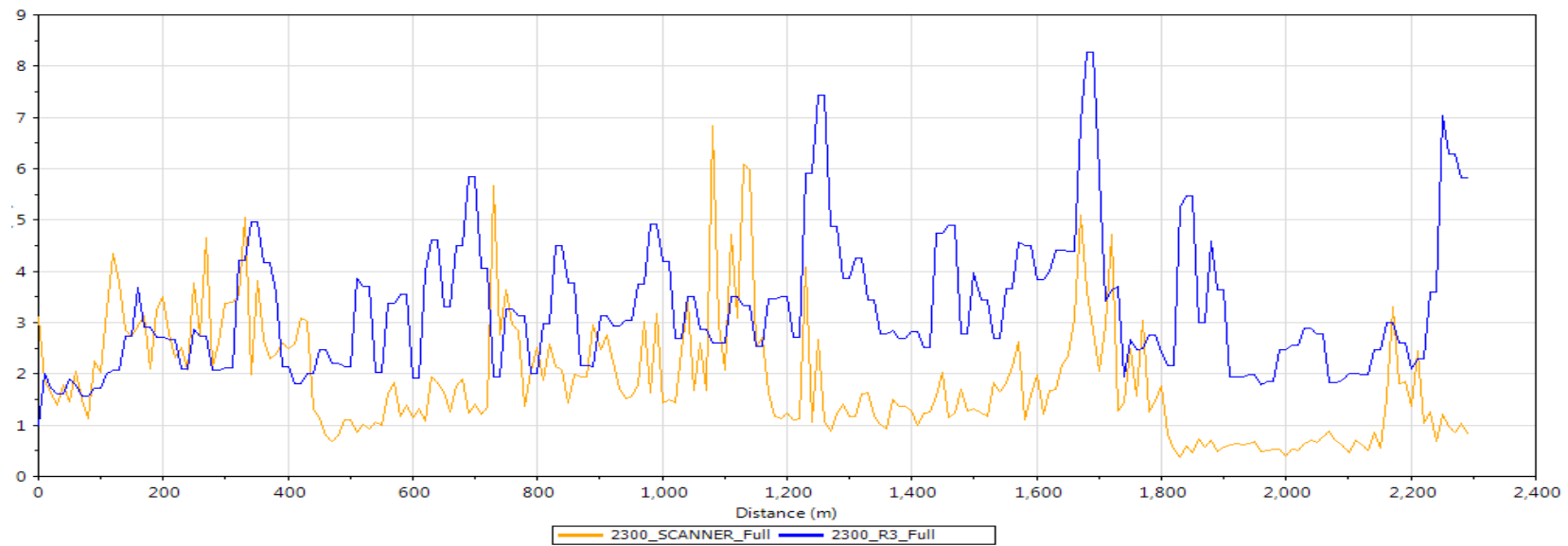


Figure C.6: Spectral Analysis of Roadroid Run 3 and SCANNER

Section Two

The second section is 1700m long and start, end points are shown below.

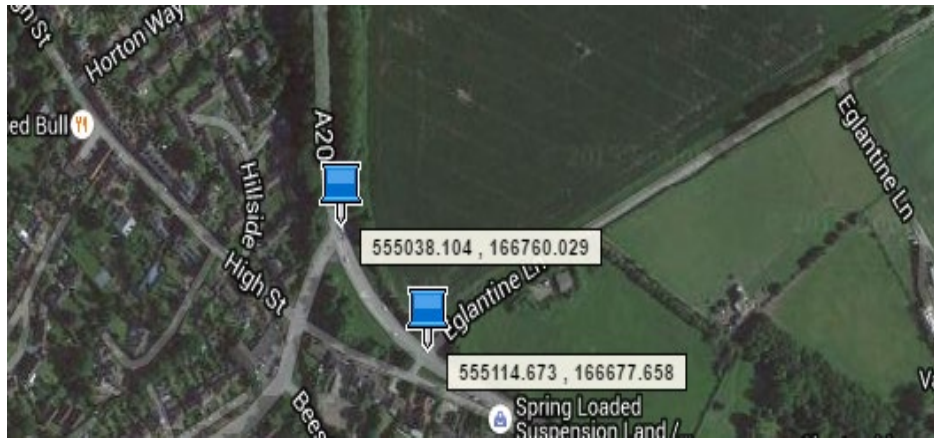


Figure C.7: The first part of road in Section Two (start, end points)

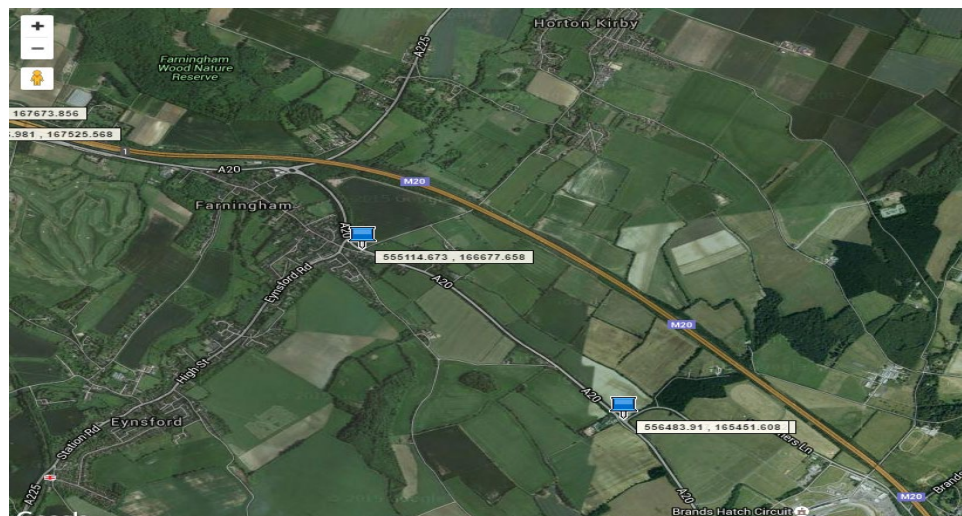


Figure C.8: The second part of road in Section Two (start, end points)

IRI comparisons in the both time domain and frequency domain as shown in the **Figures C.9 to C.11**. **Table C.2** presented the statistical analysis. By inspecting these figures and table, it can be seen that the data measured from Roadroid and SCANNER are not in agreement. Data measured by Roadroid Application generally have higher IRI values for

undulates on the rough road sections. Typically, Roadroid tend to over-estimate measured IRI.

Table C.2: The statistical analysis of IRI measurements from Roadroid and SCANNER

	SCANNER	Roadroid Application		
		Run 1	Run 2	Run 3
<i>Average measured IRI (m/km)</i>	0.883	1.676	2.618	2.985
<i>Variance (m/km)</i>	-	0.794	0.883	2.102
<i>Ratio of Error</i>	1	2.187	3.435	3.866

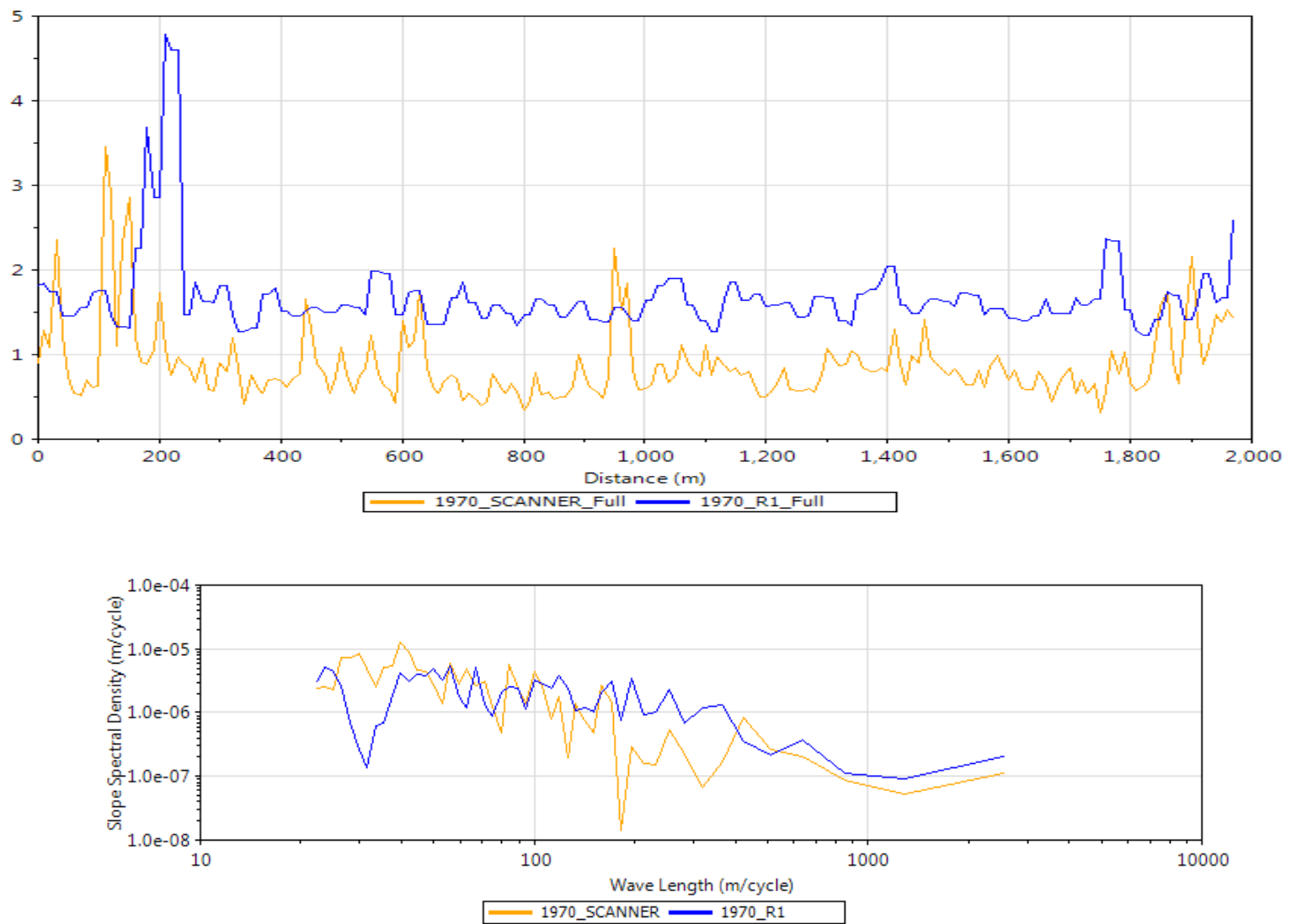


Figure C.9: Spectral Analysis of Roadroid Run 1 and SCANNER

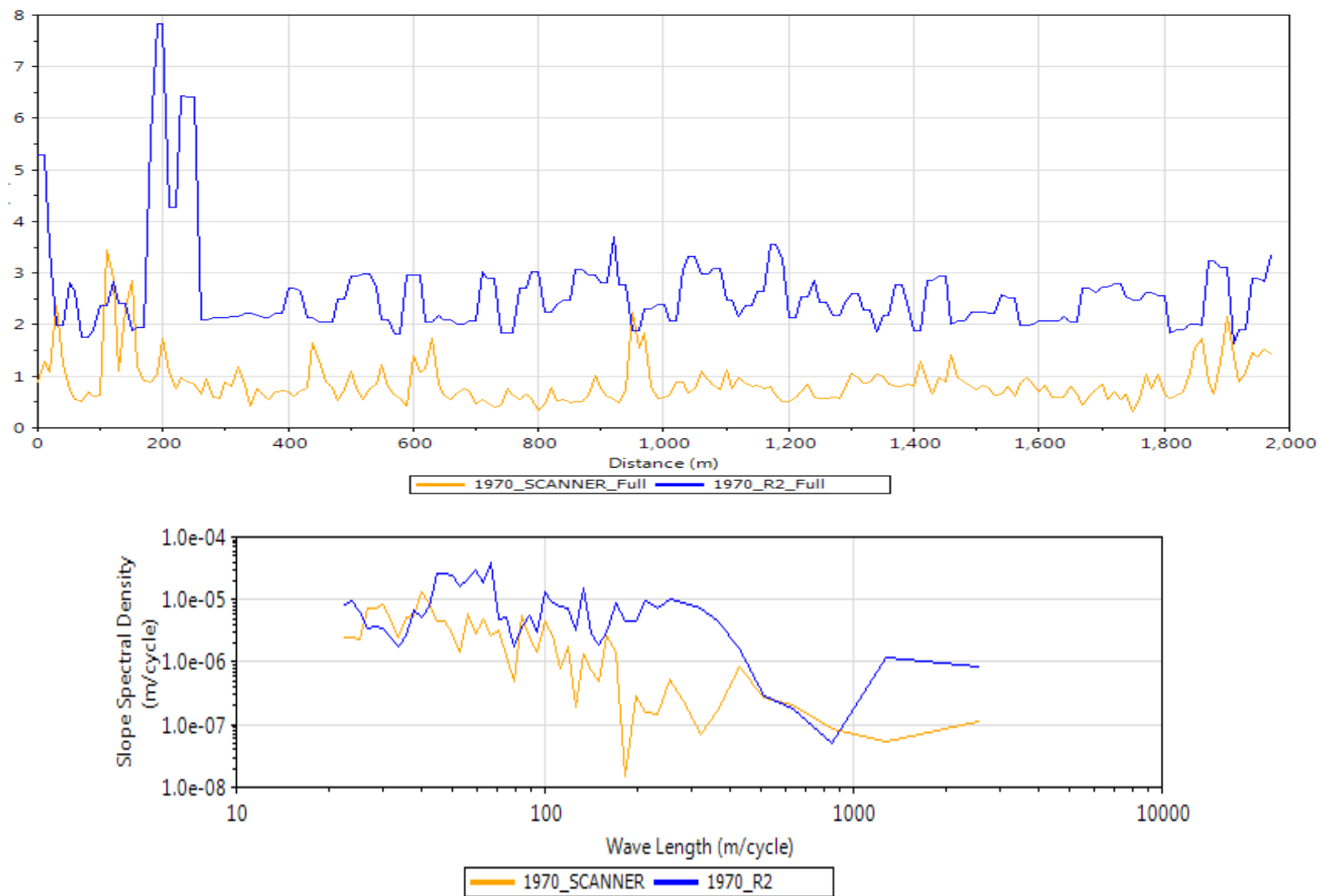


Figure C.10: Spectral Analysis of Roadroid Run 2 and SCANNER

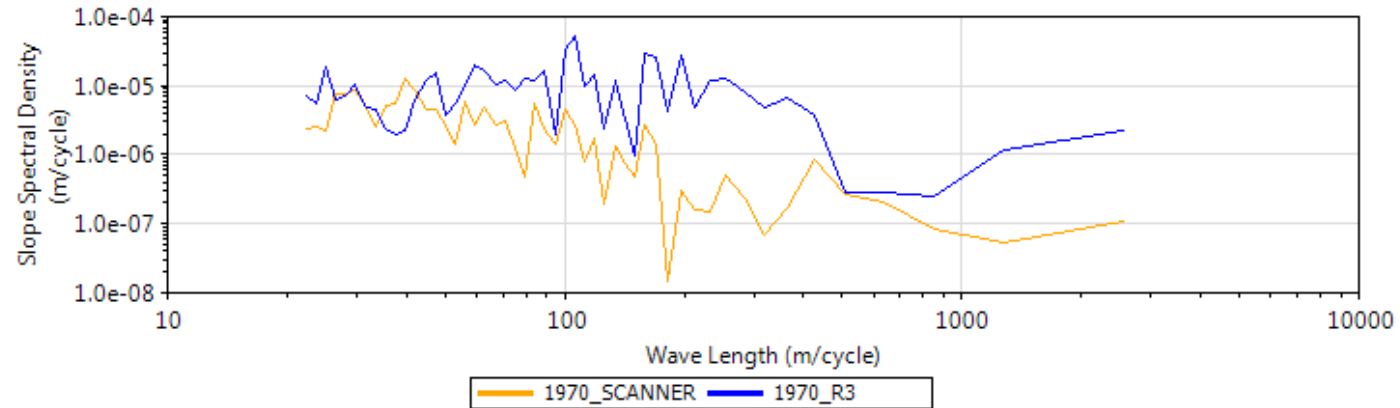
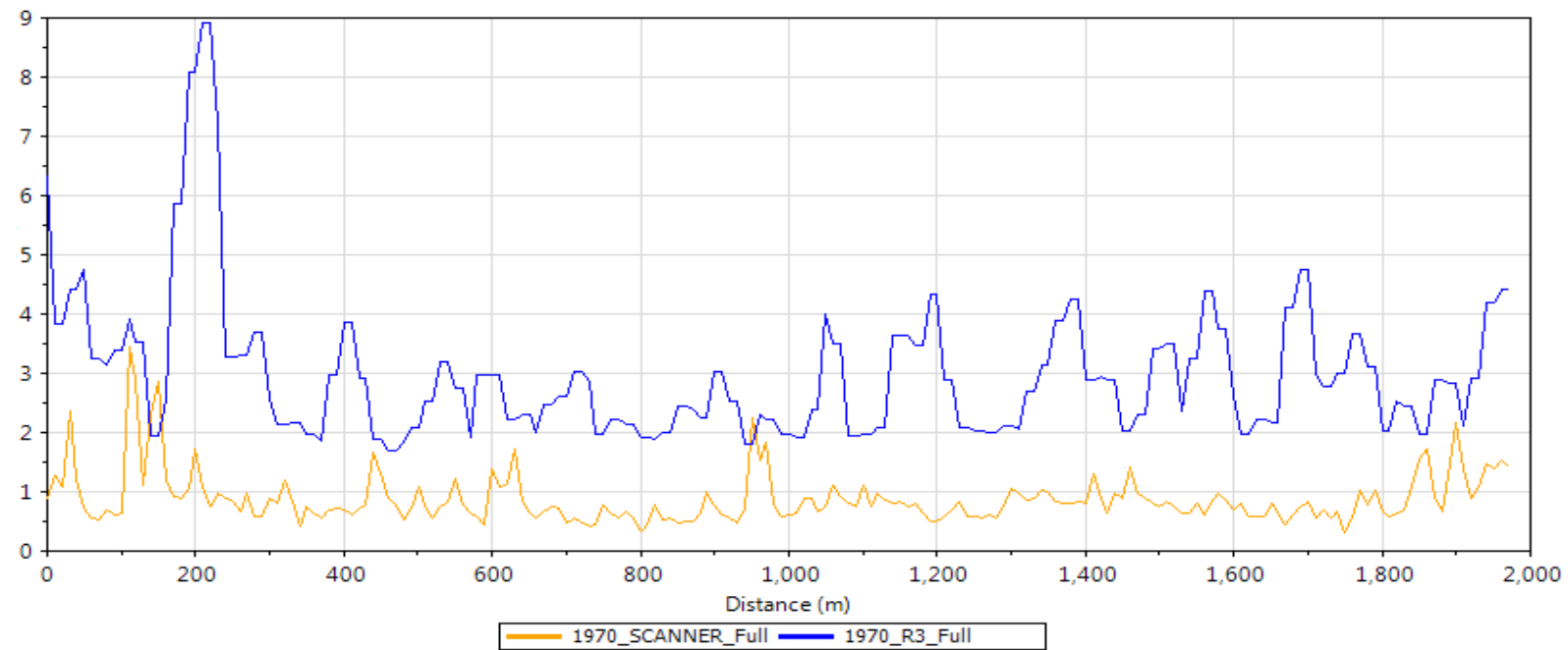


Figure C.11: Spectral Analysis of Roadroid Run 3 and SCANNER

Section Three

The third section is 1777m long and starts, end point of this section are shown in **Figure C.12** to **Figure C.13**.



Figure C.12: The first part of road in section three (start, end points)



Figure C.13: The second part of road in section three (start, end points)

IRI comparisons in the both time domain and frequency domain are shown in the **Figures C.14 to C.16**. **Table C.3** has summarized the statistical analysis between two methods.

As can be seen from the frequency domain analysis, the PSDs of Roadroid and SCANNER actually shown a level of parallel, this suggested that shapes of road roughness has been recorded correctly by Roadroid however the values measured by Roadroid were still found to be higher than SCANNER data (over-estimated). Also, bumps over chainage 800 m to 1400m were successfully detected by Roadroid Application.

Table C.3: The statistical analysis of IRI measurements from Roadroid and SCANNER

	<i>SCANNER</i>	<i>Roadroid Application</i>		
		<i>Run 1</i>	<i>Run 2</i>	<i>Run 3</i>
<i>Average Measured</i>	<i>1.23</i>	<i>2.64</i>	<i>3.76</i>	<i>4.34</i>
<i>IRI (m/km)</i>				
<i>Variance (m/km)</i>	<i>-</i>	<i>1.41</i>	<i>2.54</i>	<i>3.11</i>
<i>Ratio of Error</i>	<i>1</i>	<i>2.64</i>	<i>3.83</i>	<i>4.24</i>

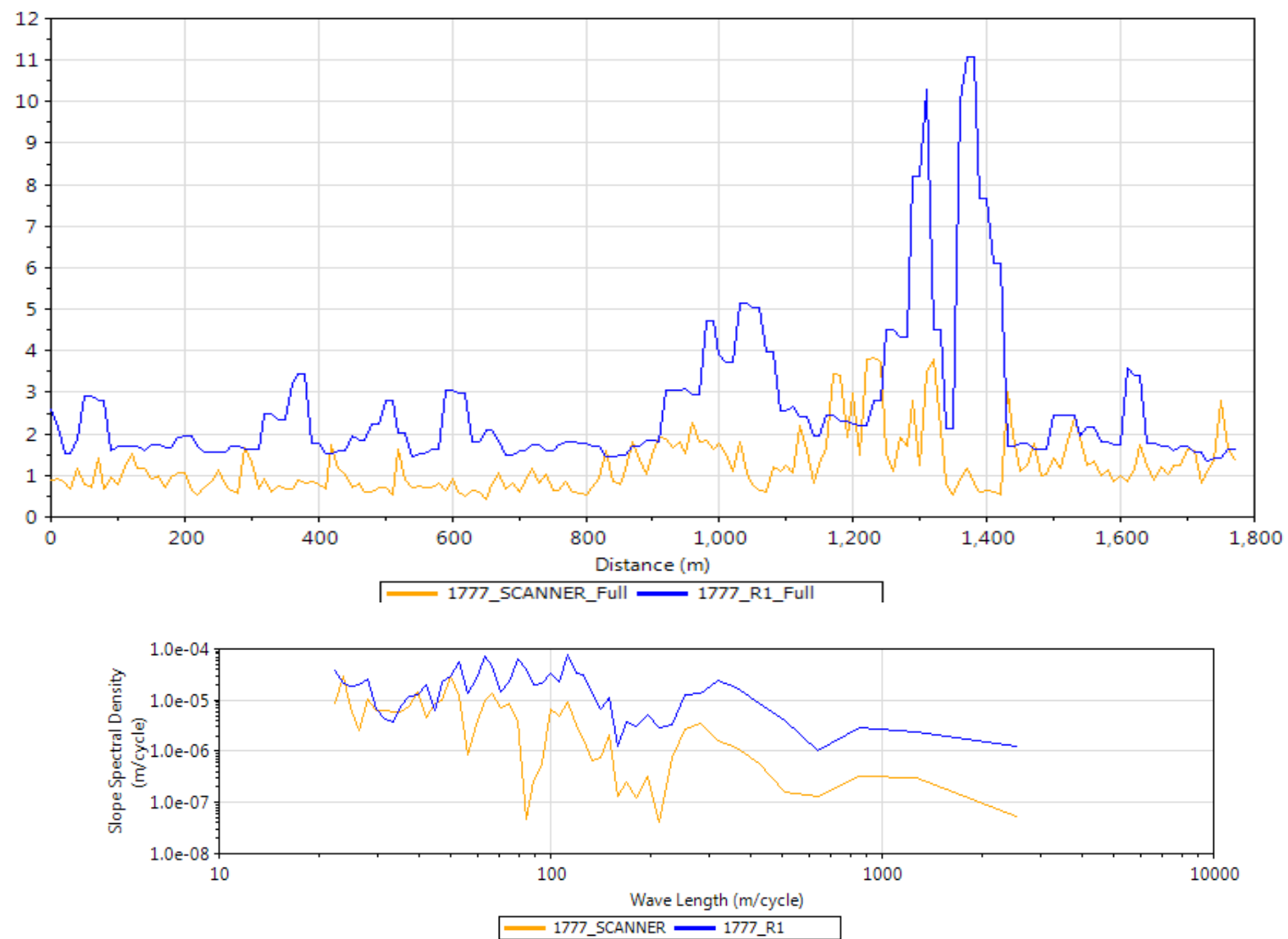


Figure C.14: Spectral Analysis of Roadroid Run 1 and SCANNER

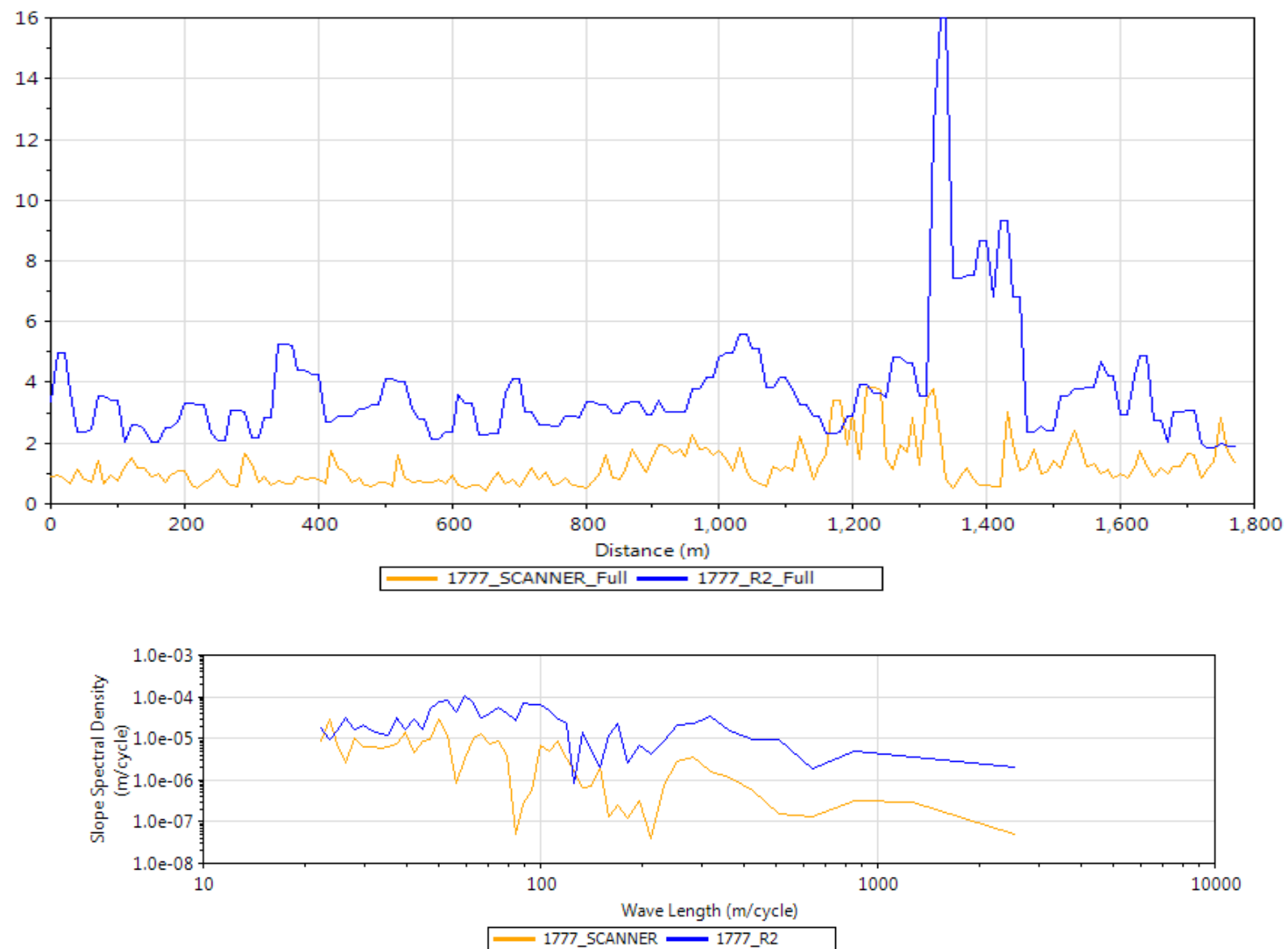


Figure C.15: Spectral Analysis of Roadroid Run 2 and SCANNER

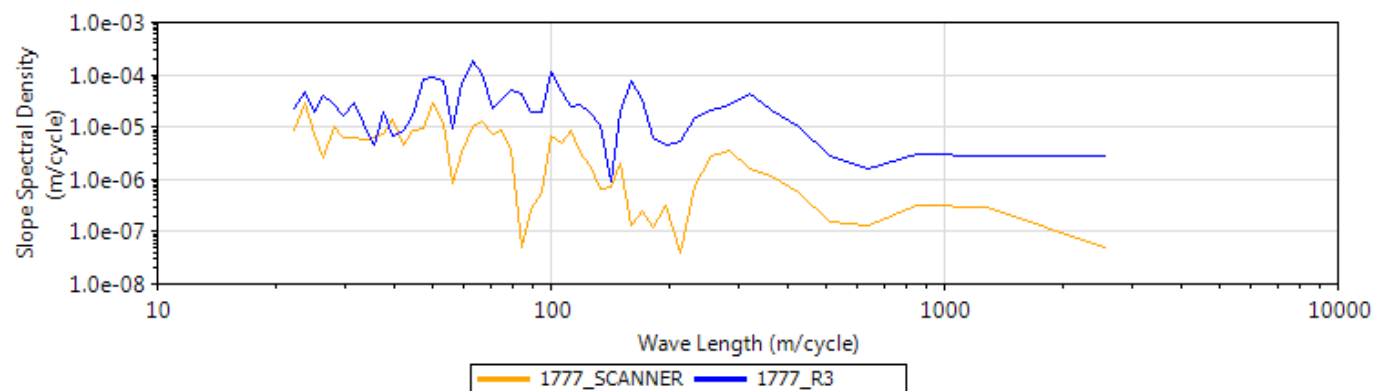
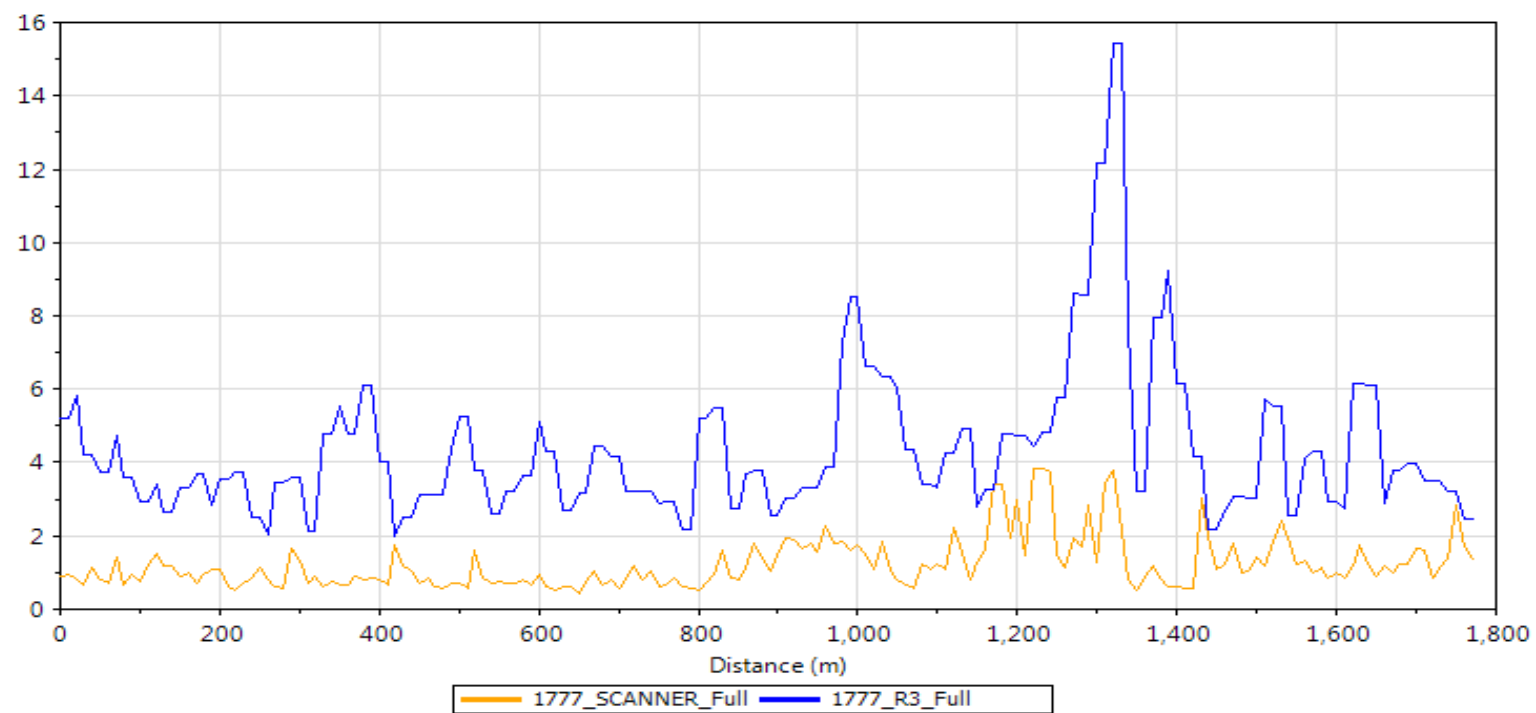


Figure C.16: Spectral Analysis of Roadroid Run 3 and SCANNER

Section Four

The fourth section is 1773m in length and start, end points were shown in **Figure C.17** to **Figure C.19**.

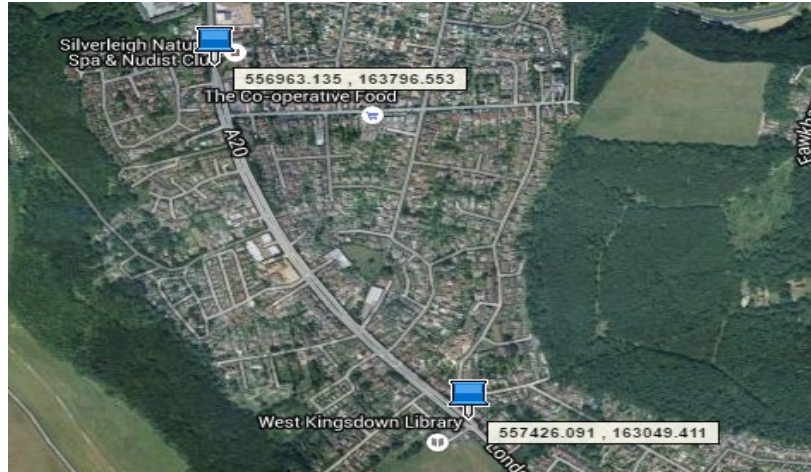


Figure C.17: The first part of road in section three (start, end points)

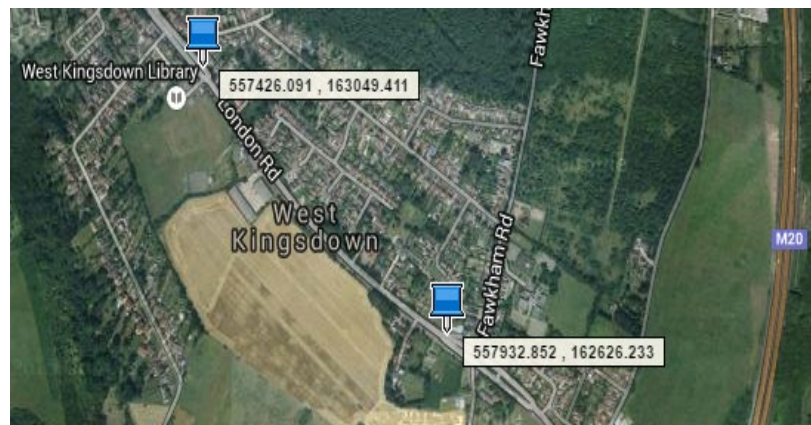


Figure C.18: The second part of road in section three (start, end points)



Figure C.19: The third part of road in section three (start, end points)

IRI comparisons in the both time domain and the frequency domain are shown in **Figures C.20- C.22**. **Table C.4** presents the statistical analysis of two methods. It can be seen that the measured datasets from Roadroid Application and SCANNER are barely related. The repeatability of the Roadroid data in this section was also appears to be poor. Roadroid has over-estimated the roughness values.

Table C.4: The statistical analysis of IRI measurements from Roadroid and SCANNER

	SCANNER	Roadroid Application		
		R1	R2	R3
Average Measured	1.10	1.92	2.86	3.26
IRI (m/km)				
Variance (m/km)	-	0.82	1.76	2.16
Ratio of Error	1	1.94	2.86	3.23

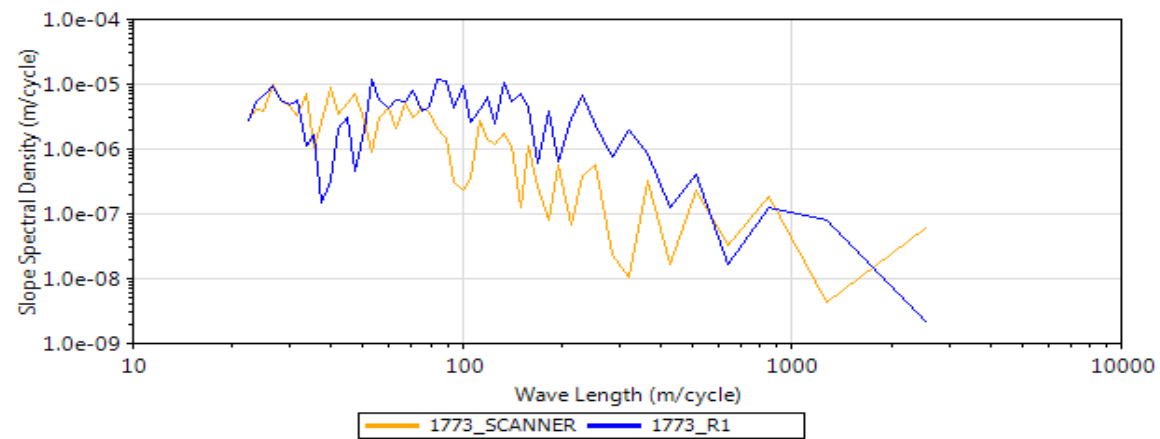
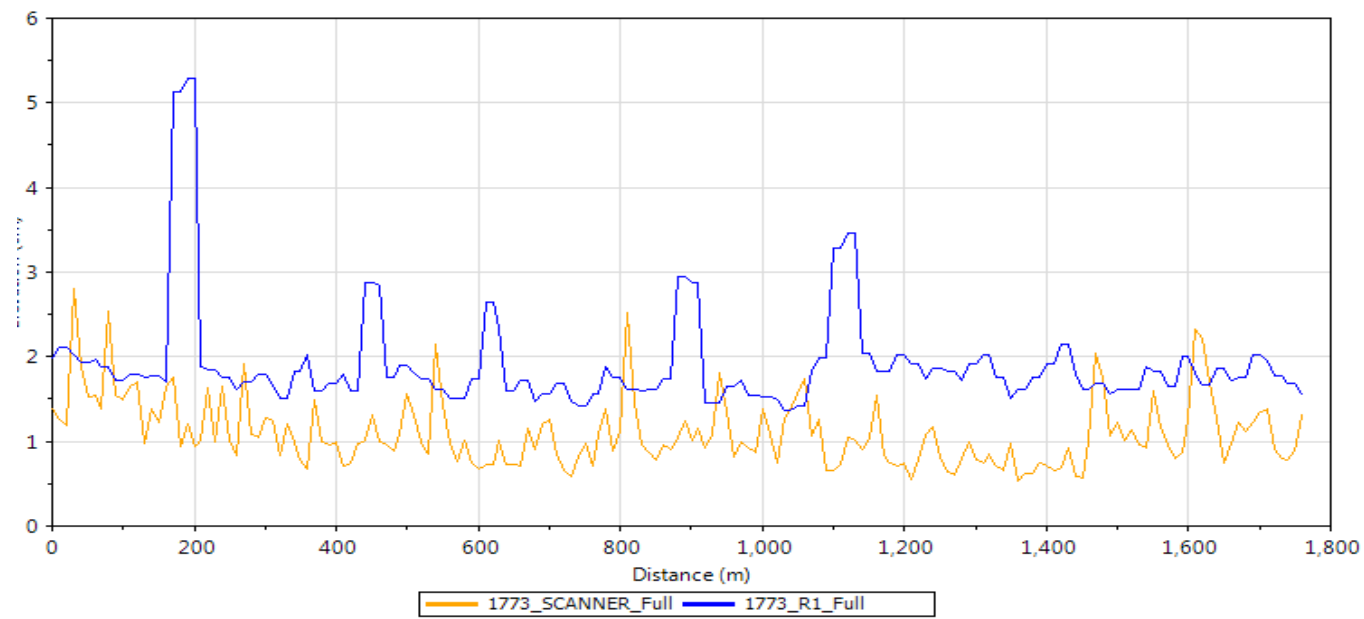


Figure C.20: Spectral Analysis of Roadroid Run 1 and SCANNER

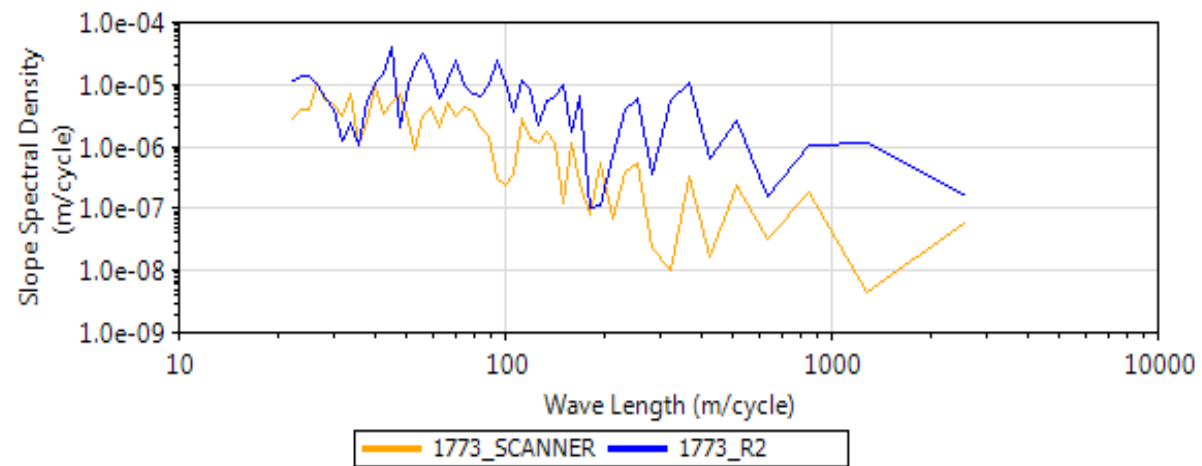
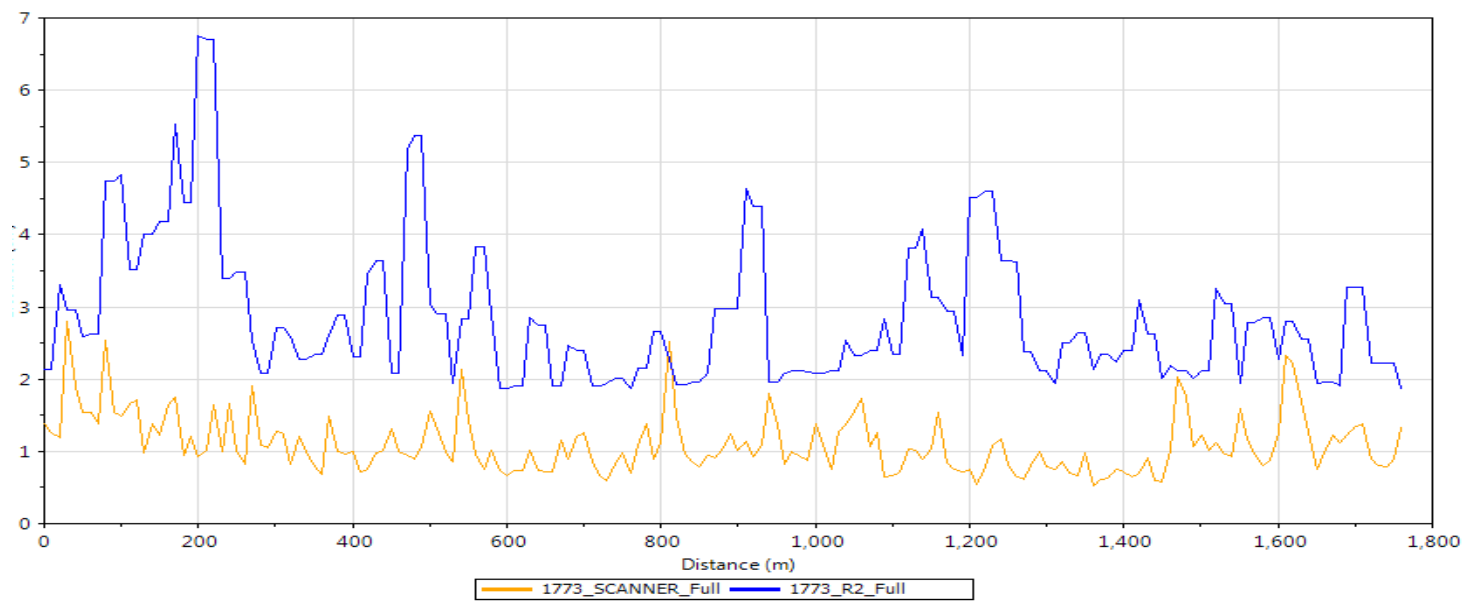


Figure C.21: Spectral Analysis of Roadroid Run 2 and SCANNER

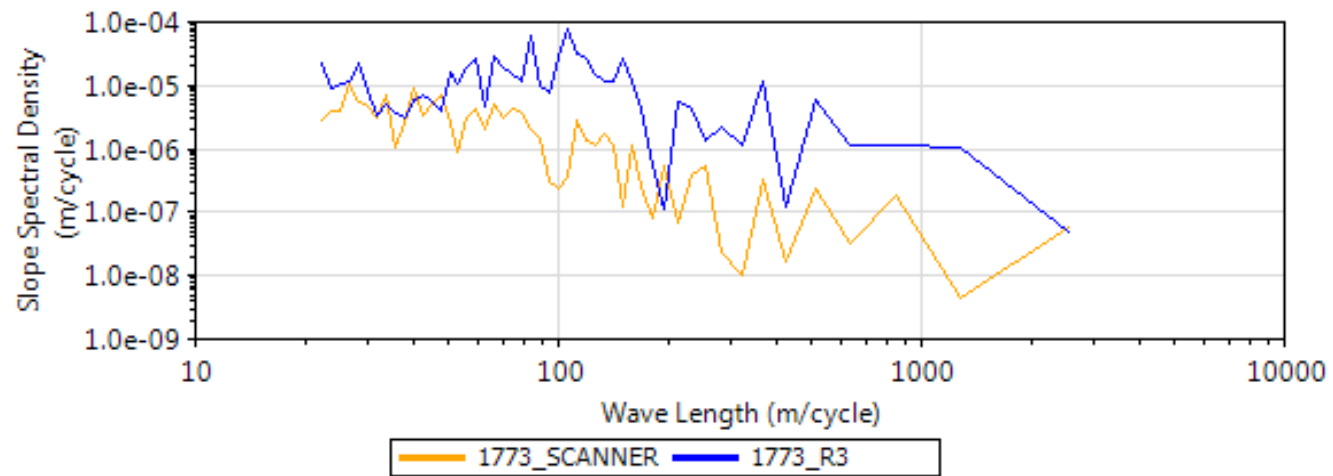
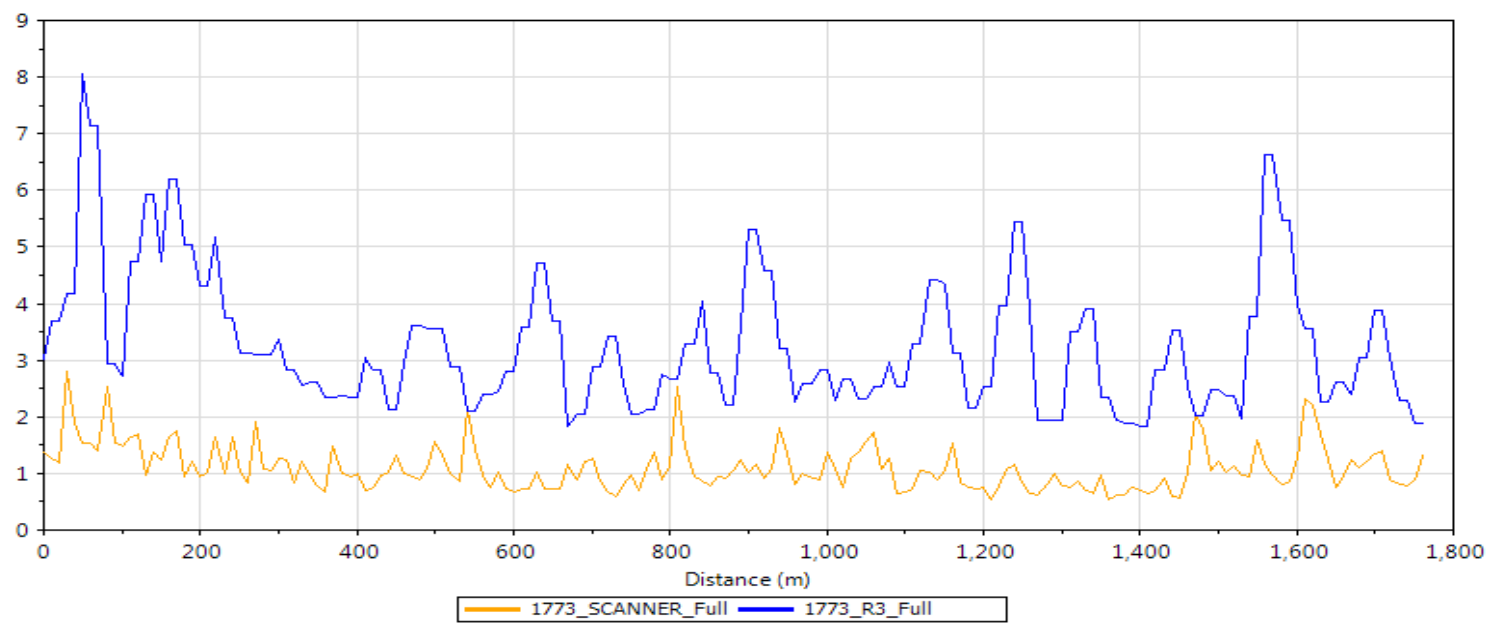


Figure C.22: Spectral Analysis of Roadroid Run 3 and SCANNER

Section Five

The fifth section is 3079m long in length and the start, end points of this sections were shown in **Figure C.23-C.24**.



Figure C.23: The first part of road in section three (start, end points)



Figure C.24: The second part of road in section three (start, end points)

IRI comparisons in both time domain and frequency domain were demonstrated in the **Figures C.25 to C.27**. **Table C.5** presented the statistical analysis of two methods.

As can be seen from these results, Roadroid Application measured higher IRI values compared to the SCANNER data, however there appears to be some agreements on the measured road roughness in terms of the undulation for these two methods (see PSD curves), while Roadroid Application was quite sensitive in this situation that overestimated the IRI. Poor repeatability was found within three runs of Roadroid datasets.

Table C.5: The statistical analysis of IRI measurements from Roadroid and SCANNER

	SCANNER	Roadroid Application		
		R1	R2	R3
Average Measured IRI (m/km)	1.667	3.055	4.502	4.399
Variance (m/km)	-	1.388	2.836	2.733
Ratio of Error	1	2.15	3.19	3.112

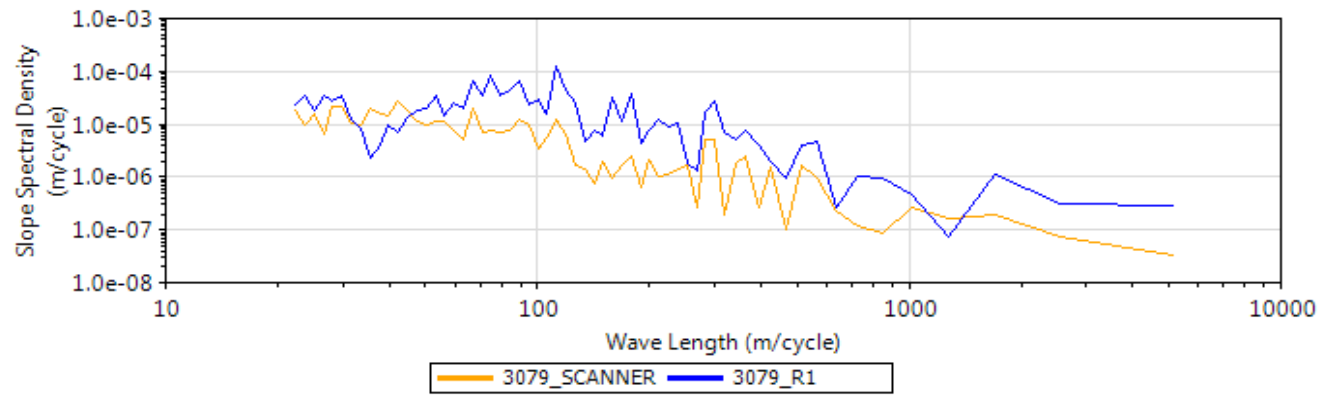
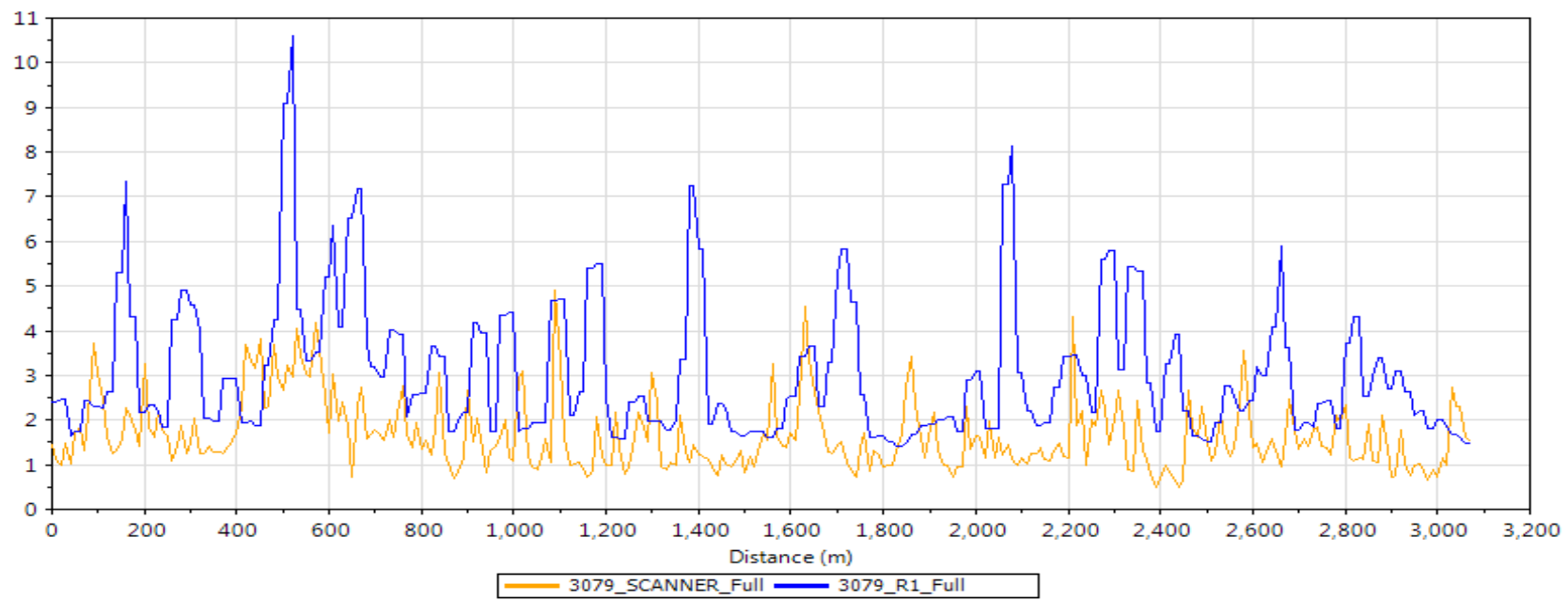


Figure C.25: Spectral Analysis of Roadroid Run 1 and SCANNER

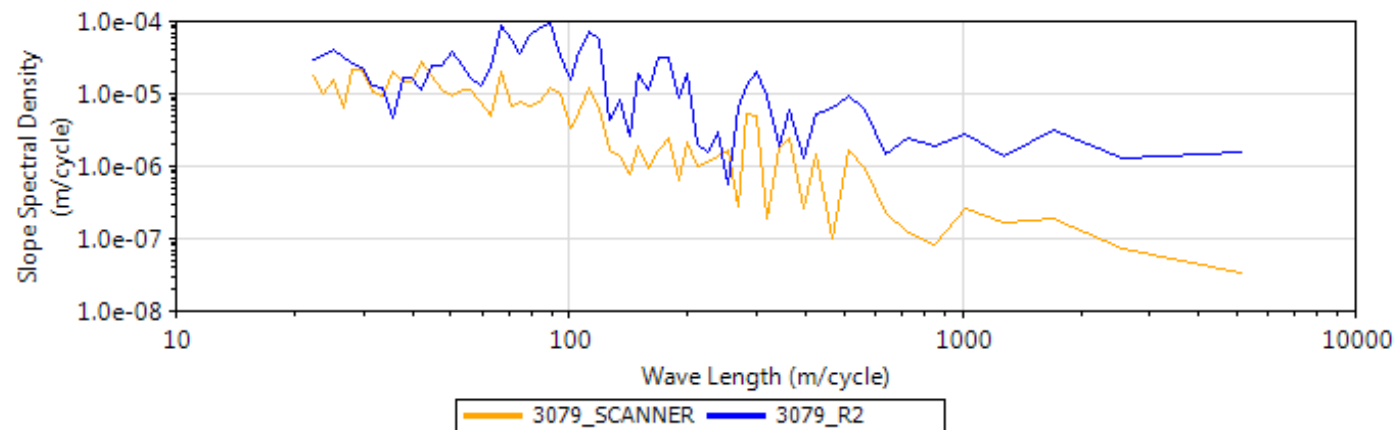
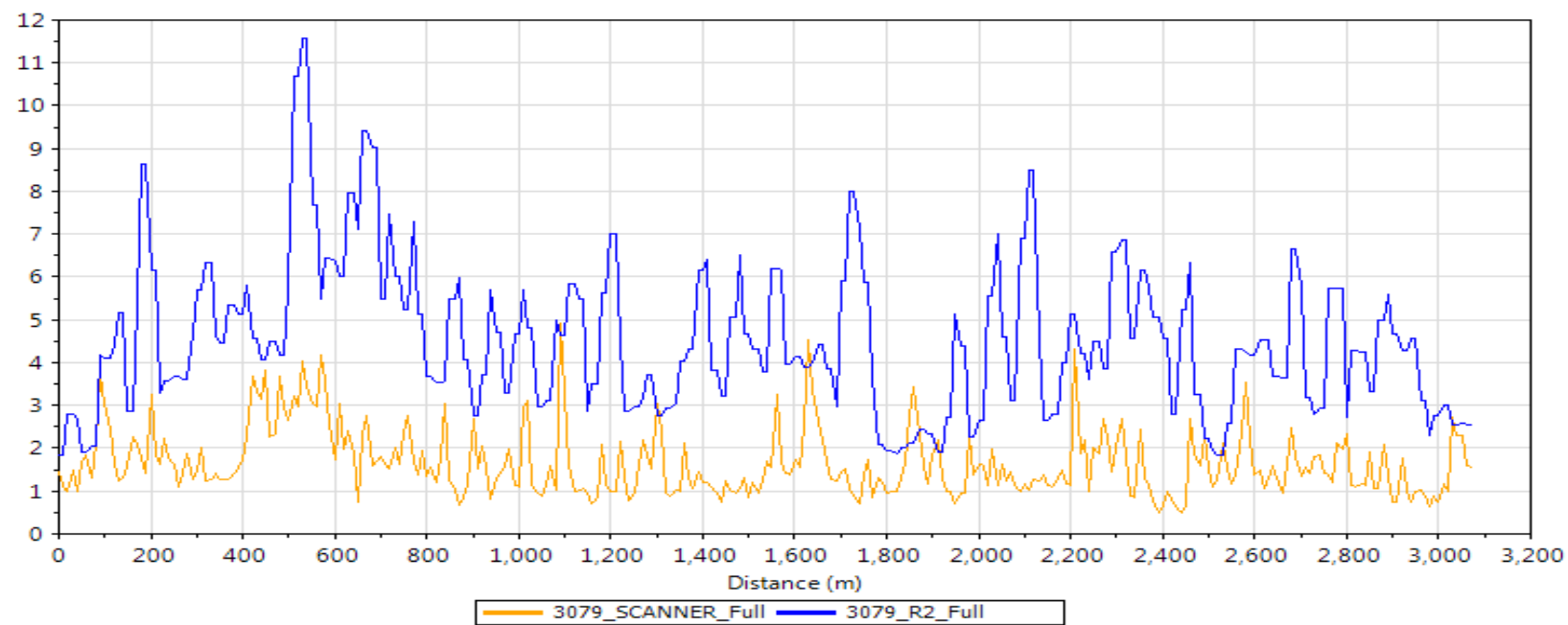


Figure C.26: Spectral Analysis of Roadroid Run 2 and SCANNER

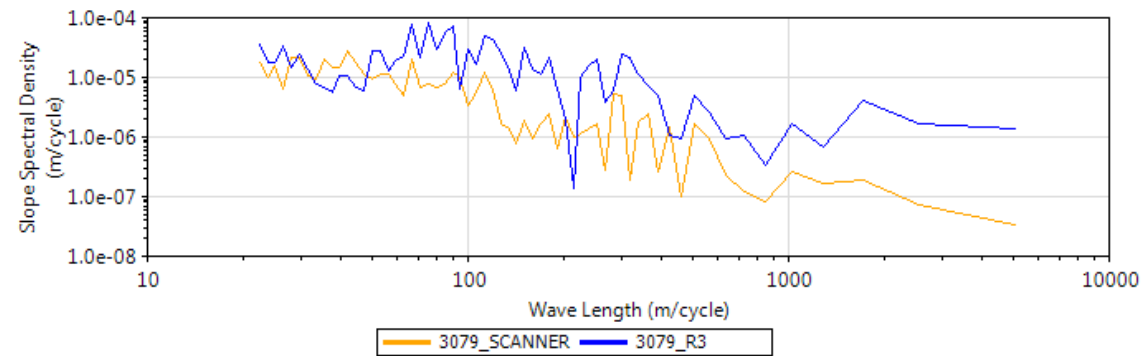
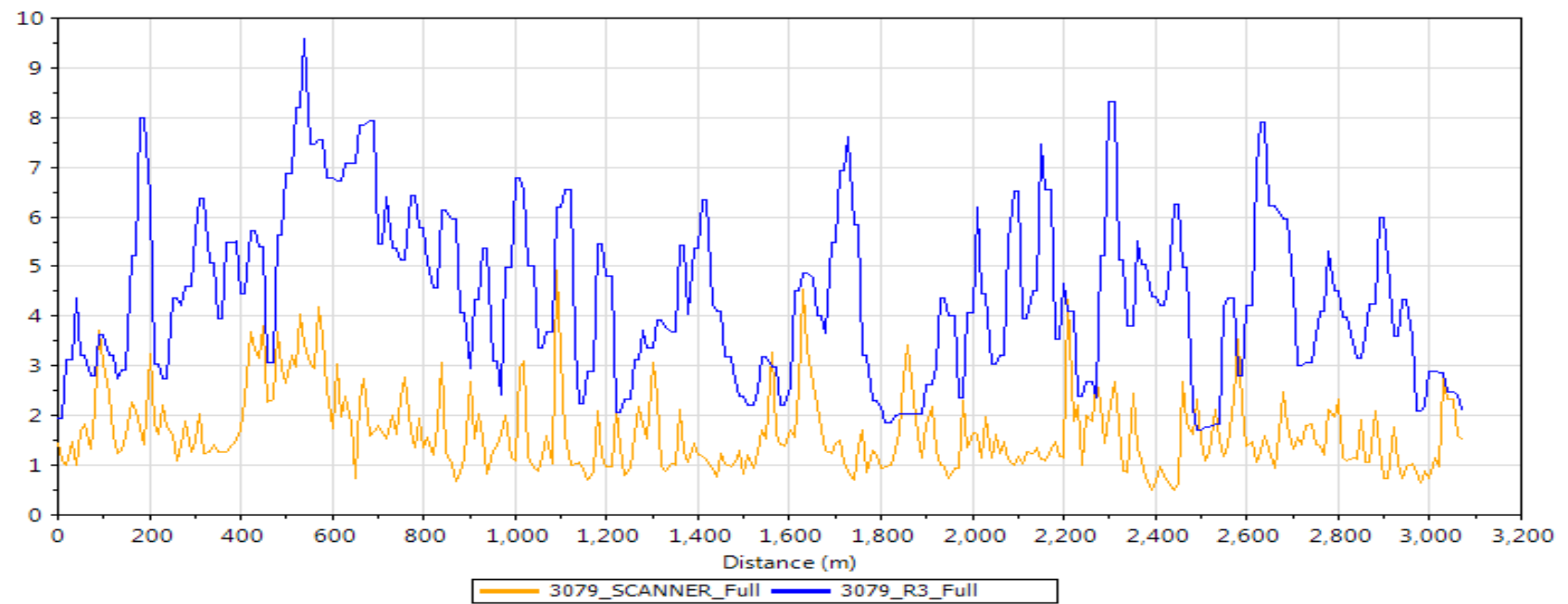


Figure C.27: Spectral Analysis of Roadroid Run 3 and SCANNER

Section Six

The last section is 950m long in length and the start, end points were shown in Figure



Figure C.28: The first part of road in section six (start, end points)



Figure C.29: The second part of road in section six (start, end points)

IRI comparisons in both time domain and frequency domain are shown in the **Figures C.30** to **C.32**. **Table C.6** summarized the statistical analysis of two methods. Results shown that Roadroid and SCANNER does not correlate well. Also, poor repeatability was found in three runs of Roadroid datasets.

Table C.6: The statistical analysis of IRI measurements from Roadroid and SCANNER

	SCANNER	Roadroid Application		
		R1	R2	R3
Average measured	1.361	2.632	4.019	3.484
IRI (m/km)				
Variance (m/km)	-	1.271	2.658	2.123
Ratio of Error	1	2.465	3.823	3.248

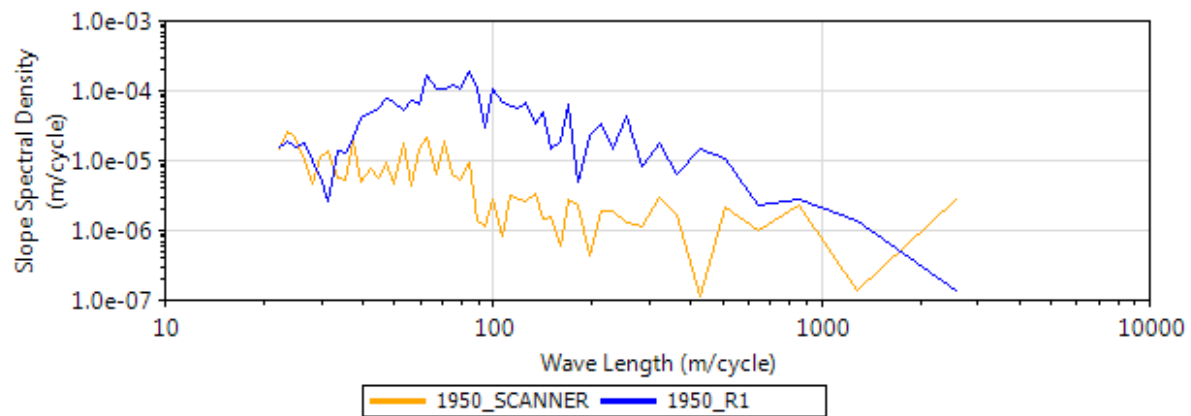
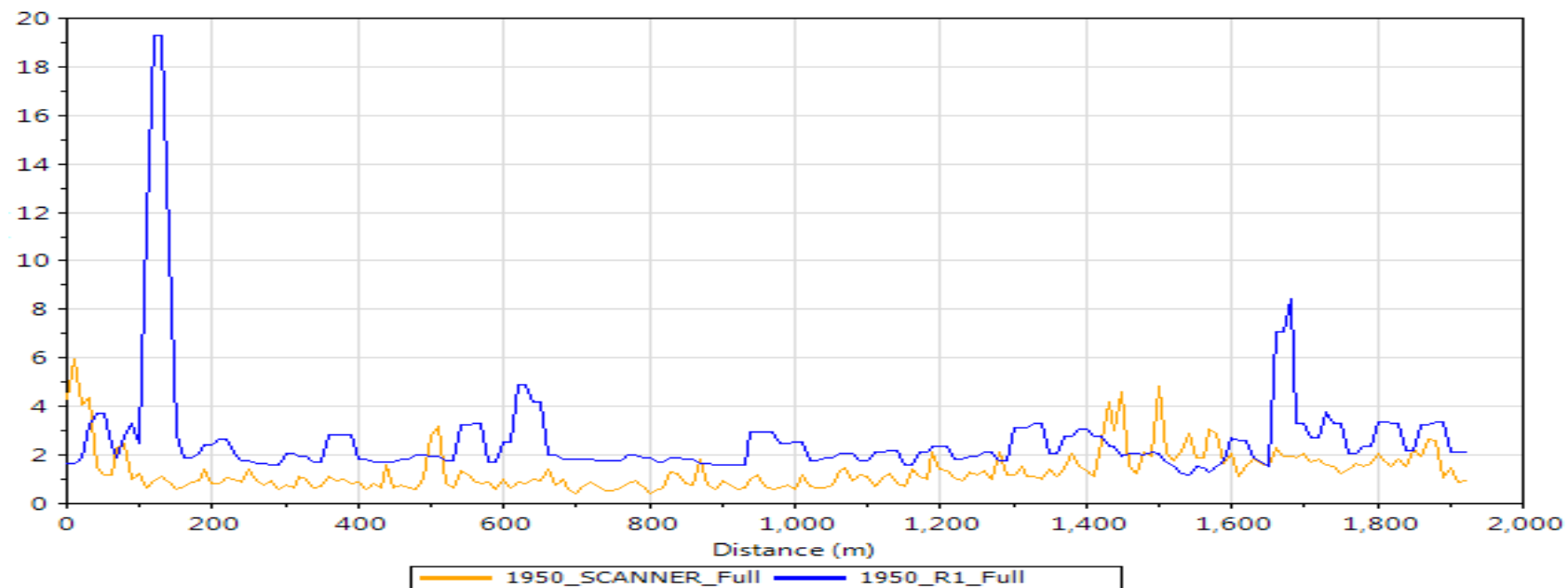


Figure C.30: Spectral Analysis of Roadroid Run 1 and SCANNER

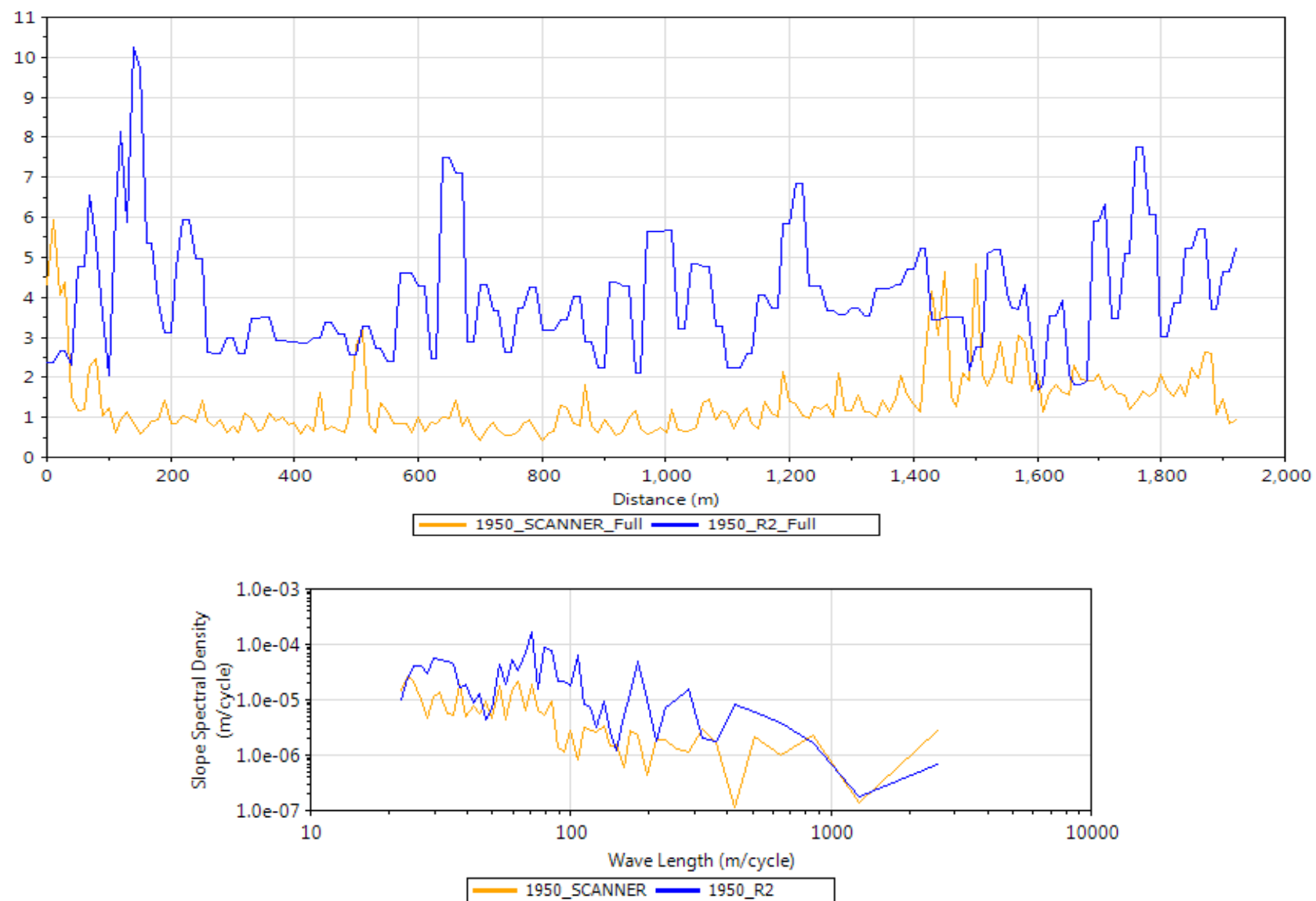


Figure C.31: Spectral Analysis of Roadroid Run 2 and SCANNER

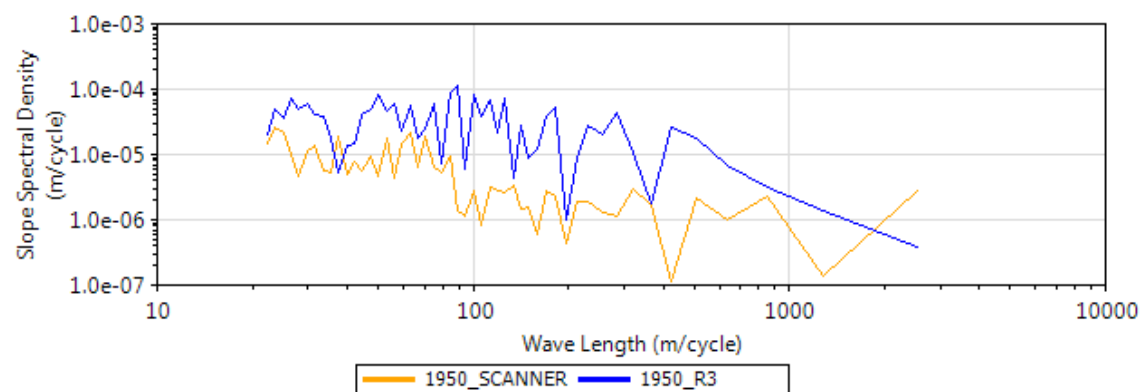
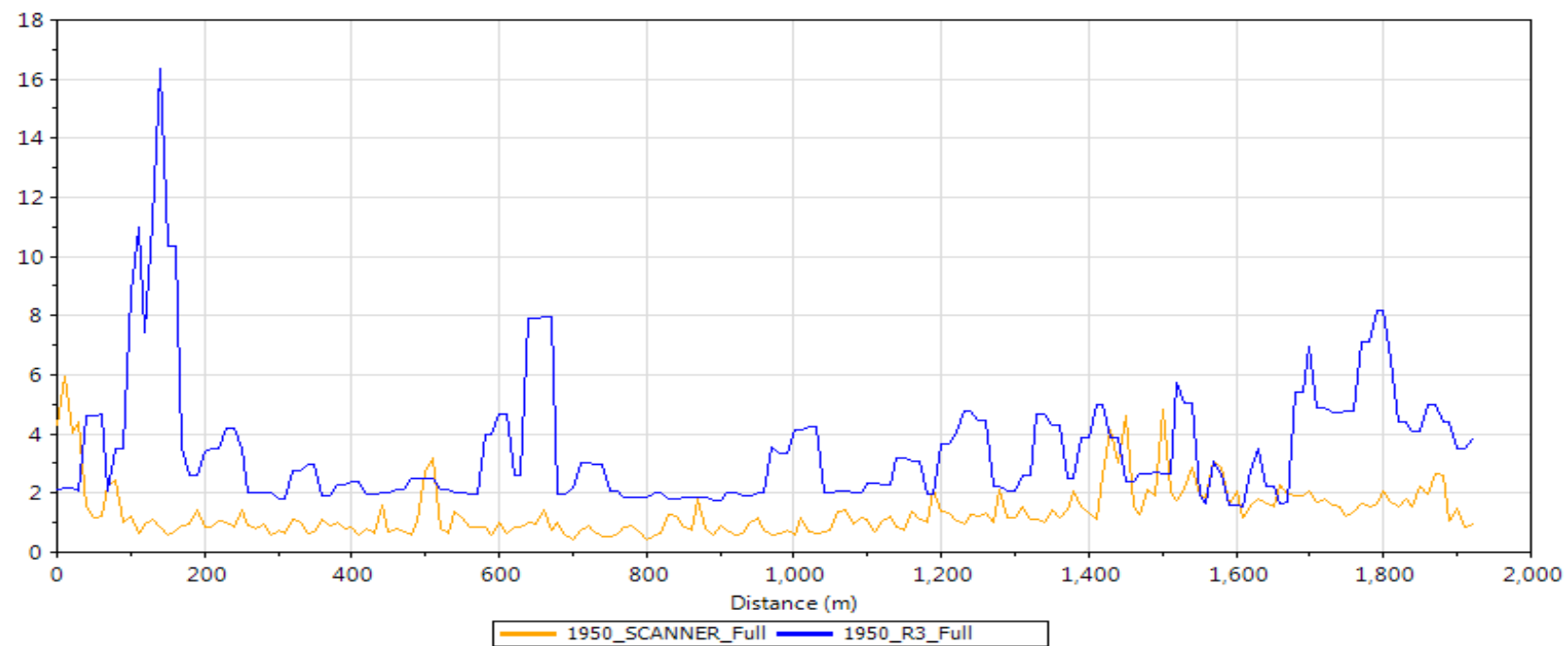


Figure C.32: Spectral Analysis of Roadroid Run 3 and SCANNER

